Geostatistical model-based predictions of helminthiases risk to assist control interventions

INAUGURALDISSERTATION
zur
Erlangung der Würde eines Doktors der Philosophie

vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von
Frédérique Chammartin
aus Villorsonnens, Freiburg, Schweiz

Basel, 2014

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von Prof. Dr. Jürg Utzinger, PD Dr. Penelope Vounatsou and Dr. Robert Bergquist.

Basel, den 18. Februar 2014

Prof. Dr. Jörg Schibler
Dekan

Just how much human helminthiasis is there in the world ? The bare mention of the question will make those of you with nosogeographical interests — or, better, helminthogeographical interests — warily scratch a mental ear and mull over a remark that ends "where angels fear to tread".

Norman R. Stoll, 1946

Summary

In the past 20 years, considerable efforts have been undertaken to fight against neglected tropical diseases, including helminthiases. The global burden of schistosomiases and soil-transmitted helminthiases due to *Ascaris lumbricoides*, *Trichuris trichiura* and hookworm is estimated to 8.4 million disability-adjusted life years (DALYs). Political awareness has emerged and specific goals and measures have been established to achieve cost-effective and sustainable reduction of the burden. Financial resources marshalled by philanthropists have permitted to advance scientific research and anthelmintic donations provide a real opportunity to achieve a significant decrease of the hardship. An in-depth knowledge and understanding of the geographical distribution of the infections is crucial for guiding and implementing control programmes. High resolution risk-estimates are not only essential for a rational and cost-effective planning of interventions, but also at later stages for evaluation and surveillance. Statistical models can be used to build a mathematical relationship between observations and risk factors and to predict the risk at unobserved locations. The propagation of helminths is governed by complex interactions of environmental and socioeconomic factors heterogeneously distributed in space. Furthermore, the presence of parasites at a given location has an influence on their occurrence in the surroundings. Thus, observations are spatially correlated and risk modelling should be handled within a spatial statistical approach. Bayesian geostatistical framework offers a flexible and rigorous methodology to deal with such data, while allowing the model to take into account spatial uncertainty.

Chapter 1 introduces some fundamentals related to this PhD thesis which aims to provide geostatistical model-based predictions of helminthiases risk to assist control interventions. An overview of the biology and epidemiology of the schistosomes and the soil-transmitted helminths studied is offered and treatments and diagnostic tools currently available are addressed. Strategies and objectives of the control programmes, as well as the basics of the Bayesian geostatistical methodology for infection risk modelling are discussed.

The overall goal of this work is to develop, implement, and validate Bayesian geostatistical methodology within an epidemiological setting. Focus is placed on the estimation of the geographical distribution of helminthiases risk, the ultimate aim being to fill both methodological and epidemiological gaps that have to be addressed to respond to the urgent need to identify spatial targets for interventions.

Important issues related to the spatial modelling of infectious disease risk, including Bayesian computation of large datasets, heterogeneity of historical survey data, stationary and isotropy assumptions, and variable selection approaches are reviewed in Chapter 2. Emphasis is placed on Bayesian geostatistical variable selection with a parameterisation of the regression coefficients through a parameter expanded normal mixture of the inverse-gamma (peNMIG) distribution, which allows selection of blocks of covariates, particularly categorical variables. The method is illustrated on historical prevalence data of *Schistosoma mansoni* in Côte d'Ivoire and is compared to more traditional selection techniques. This innovative parameterization provides a rigorous approach for the selection of predictors within a Bayesian geostatistical framework (especially in the presence of categorical variables), identifies the most important predictors of *S. mansoni* infection risk and leads to a more parsimonious model compared to traditional selection approaches that ignore the spatial structure in the data.

In Chapter 3, we propose an analysis of school-aged children infection status for *S. mansoni* and *S. haematobium* that have been collected during a national cross-sectional survey, conducted from November 2011 to February 2012 in Côte d'Ivoire. A Bayesian geostatistical multinomial model is implemented to estimate each mono-infection risk, as well as co-infection. The overall schistosomiasis risk among school-aged children is estimated to be 8.9% (5.3% with *S. haematobium* and 3.8% with *S. mansoni*) and approximately 2 million treatments would be necessary at health district level to control schistosomiasis with deworming drugs. The distinct spatial patterns of *S. haematobium* and *S. mansoni* imply that co-infection with these two types of parasite is low across the country.

Chapter 4 presents the results of a systematic review and a geostatistical meta-analysis of prevalence survey data pertaining on soil-transmitted helminth infections in South America. Out of 4 085 scientific papers examined, 174 have provided relevant data. Observations are sparse for the south and the western coast of the subcontinent. No relevant information has been identified for Uruguay and little data have been collected for smaller countries such as Suriname, Guyana, French Guiana, and Ecuador. In total, ascariasis, trichuriasis

and hookworm infection prevalences have been extracted for 6 948 locations. Large datasets have led us to develop Bayesian geostatistical models where spatial uncertainty is estimated through the predictive process approximation. Risk estimations across South America offer important baseline support for spatial targeting of control measures and suggest that surveys informing about the soil-transmitted helminthiases situation are needed, especially in countries where current data is scarce and the estimated risk, driven by climatic suitability for parasites transmission, is high. In South America, the risk of contracting soil-transmitted helminthiases has substantially reduced since 2005 and the population-adjusted prevalence is estimated to be 15.6%, 12.5% and 11.9% for ascariasis, trichuriasis and hookworm infection, respectively.

In Chapter 5, the peNMIG Bayesian geostatistical variable selection has been further developed to allow selection of no more than one predictor among groups of highly correlated covariates. This approach has identified the most important risk factors to build Bayesian geostatistical models of the three major soil-transmitted helminths infection risks in Bolivia. High resolution risk estimates indicate that 48.4% of the Bolivian population is infected with any soil-transmitted helminth and 2.9 million annualised treatments would be required for preventive chemotherapy. However, the scarcity of the data suggests that a national survey is required for more accurate mapping that will govern spatial targeting of control interventions.

On the basis of historical data from 1995 onwards, spatio-temporal distribution of the three major soil-transmitted helminthiases in Brazil is analysed in Chapter 6. Our spatio-temporal models consider a space-time interaction, allowing the spatial structure to vary across time. Models are built on environmental and socioeconomic predictors selected through a Bayesian geostatistical variable selection for large datasets, that addresses non-linearity and correlation of the explanatory variables. Spatio-temporal models are fitted within the integrated nested Laplace approximation/stochastic partial differential equations framework. We show that the risk of *A. lumbricoides* and hookworm infections has decreased over the past 20 years in Brazil. From 2010 onwards, the risk is estimated to 3.6% for *A. lumbricoides*, 1.7% for hookworm, and 1.4% for *T. trichiura*. Thus, the number of annualised treatments required for school-aged children mass deworming at municipality level is estimated to 1.8 million.

The work presented in this thesis offers Bayesian statistical methodology for spatial modelling and estimation of helminthiases risk. We propose and implement a Bayesian

geostatistical variable selection that addresses non-linearity and correlation of the predictors, as well as approaches for spatio-temporal and large datasets modelling. Our main contribution to the field of helminth epidemiology consists in informing national control programmes with high spatial resolution risk estimates and number of people in need of preventive chemotherapy to control soil-transmitted helminth infections in South America and schistosomiasis Côte d'Ivoire. Thus, our work provides an important benchmark on which further estimates could be derived, as soon as new data become available and control interventions are progressing.

Zusammenfassung

In den vergangenen 20 Jahren wurde im Kampf gegen vernachlässigte tropische Krankheiten, einschließlich Wurmerkrankungen, ein enormer Aufwand unternommen. Schätzungen zufolge liegt die globale Belastung, welche von Schistosomiasis und durch Bodenkontakt übertragene Wurmerkrankungen aufgrund von *Ascaris lumbricoides*, *Trichuris trichiura* und Hakenwürmer verursacht wird, bei 8.4 Millionen disability-adjusted life years (DALYs). Das politische Bewusstsein wurde gefördert und spezielle Ziele und Maßnahmen entwickelt, um einen kosteneffektiven und nachhaltigen Rückgang der Last zu erzielen. Durch Philanthropen mobilisierte finanzielle Ressourcen haben es ermöglicht die wissenschaftliche Forschung anzutreiben. Weiter bieten Spenden von Medikamenten gegen Wurminfektionen eine reelle Chance für einen signifikanten Rückgang des Elends. Fundiertes Wissen und Verständnis der geographischen Verteilung der Infektionen sind unabdingbar für die Steuerung und Implementierung von Kontrollprogrammen. Risikoschätzungen mit hoher räumlicher Auflösung sind essentiell für eine rationale und kosteneffektive Interventionsplanung. Darüber hinaus spielen sie eine wichtige Rolle im Rahmen der Evaluierung und Überwachung. Statistische Modelle können dazu dienen eine mathematische Beziehung zwischen Beobachtungen und Risikofaktoren herzustellen und das Risiko an Orten ohne Daten vorherzusagen. Die Ausbreitung der Würmer wird durch komplexe Interaktionen von räumlich heterogen-verteilten Umwelt- und sozioökonomischen Faktoren beherrscht. Des Weiteren beeinflusst das Parasitenvorkommen an einem Ort das Auftreten jener in der Umgebung. Beobachtungen sind daher räumlich korreliert und die Modellierung des Risikos sollte innerhalb eines Ansatzes der räumlichen Statistik durchgeführt werden. Der Bayes'sche geostatistische Ansatz bietet eine flexible und geeignete Methodologie, um mit dieser Art von Daten umzugehen, indem das Model räumliche Unsicherheit berücksichtigt.

Kapitel 1 führt einige Grundlagen für diese PhD Thesis ein, welche geostatistische modellbasierte Vorhersagen des Wurmerkrankungsrisikos bereitstellt, um Kontrollinterventionen zu unterstützen. Neben der Biologie und Epidemiologie der Schistosomen und der in

dieser Arbeit behandelten durch Bodenkontakt übertragenen Würmer werden gegenwärtig verfügbare Behandlungsmöglichkeiten und diagnostische Methoden präsentiert. Darüber hinaus werden Strategien und Zielsetzungen der Kontrollprogramme sowie die Grundlagen der Bayes'schen geostatistischen Methodologie im Rahmen der Modellierung des Infektionsrisikos besprochen. Das Gesamtziel dieser Arbeit ist die Entwicklung, Implementierung und Validierung von Bayes'schen geostatistischen Methoden innerhalb eines epidemiologischen Kontextes. Die Schätzung der geographischen Verteilung des Wurmerkrankungsrisikos steht hierbei im Mittelpunkt, wobei das oberste Ziel das Schließen jener methodologischen und epidemiologischen Lücken ist, welche die dringend erforderliche Ermittlung von Interventionszielorten ermöglichen.

Kapitel 2 behandelt wichtige Themen bezüglich räumlicher Modellierung des Infektionskrankheitsrisikos wie zum Beispiel Bayes'sche Berechnung von großen Datensätzen, Heterogenität historischer Erhebungsdaten, Annahmen der Stationarität und Isotropie sowie Variablenauswahlverfahren. Im Fokus steht die Bayes'sche geostatistische Variablenauswahl mit einer Parametrisierung der Regressionskoeffizienten mittels parametrisch-erweiterten Normal-Mischverteilung der inversen Gammaverteilung. Jener Ansatz ermöglicht die Auswahl von Blöcken der Kovariaten, insbesondere kategorischer Variablen. Diese Methode wird anhand historischer Prävalenzdaten für *Schistosoma mansoni* in der Elfenbeinküste verdeutlicht und mit traditionelleren Auswahlverfahren verglichen. Diese innovative Parametrisierung bietet einen angemessenen Ansatz für die Auswahl von Einflussvariablen innerhalb eines Bayes'schen geostatistischen Kontextes (insbesondere beim Vorliegen von kategorischen Variablen), bestimmt die wichtigsten Prädiktoren für das *S. mansoni* Infektionsrisiko und führt zu einem reduziertem Model im Vergleich zu üblichen Ansätzen, welche die räumliche Struktur der Daten ignorieren.

In Kapitel 3 wird eine Analyse des Infektionsstatus bezüglich *S. mansoni* und *S. haematobium* von Kindern im Schulalter vorgestellt. Die Daten wurden im Rahmen einer nationalen Querschnittserhebung gesammelt, welche von November 2011 bis Februar 2012 in der Elfenbeinküste durchgeführt wurde. Hierzu wurde ein Bayes'sches geostatistisches multinomiales Model eingesetzt, welches sowohl das Monoinfektionsrisiko als auch das Koinfektionsrisiko schätzt. Das gesamte Schistosomiasis-Risiko unter den Kindern im Schulalter wurde auf 8.9% geschätzt (5.3% mit *S. haematobium* und 3.8% mit *S. mansoni*) und ungefähr 2 Millionen Behandlungen wären auf Bezirksebene, charakterisiert durch Zugehörigkeit zu Gesundheitszentren, nötig, um Schistosomiasis mit Entwurmungsmedikamenten

zu bekämpfen. Die unterschiedlichen räumlichen Muster von *S. haematobium* und *S. mansoni* impliziert, dass die Koinfektion der beiden Parasitenarten übers Land verteilt gering ist.

In Kapitel 4 werden die Ergebnisse eines systematischen Reviews und eine geostatistische Metaanalyse von Prävalenzerhebungsdaten hinsichtlich durch Bodenkontakt übertragener Wurminfektionen in Südamerika präsentiert. Die Untersuchung umfasste 4 085 wissenschaftliche Veröffentlichungen, aus denen 174 relevante Daten lieferten. Für die Süd- und Westküste des Subkontinents liegen wenige Beobachtungen vor und keinerlei relevanten Informationen wurden für Uruguay gefunden. Wenige Daten wurden für kleinere Länder wie Surinam, Guyana, Französisch-Guyana und Ecuador gesammelt. Insgesamt wurden Prävalenzen bezüglich Askariasis, Trichuriasis und Hakenwurminfektion für 6 948 Standorte extrahiert. Große Datensätze haben uns dazu veranlasst Bayes'sche geographische Modelle zu entwickeln, welche die räumliche Unsicherheit mittels der sogenannten predictive process approximation schätzen. Risikoschätzungen in ganz Südamerika bieten eine wichtige Grundlage zur Unterstützung der räumlichen Zielsetzung von Kontrollmaßnahmen. Des Weiteren verdeutlichen sie den Bedarf an Erhebungen, welche Informationen bezüglich durch Bodenkontakt übertragene Wurmerkrankungen liefern. Dies gilt insbesondere für Länder in denen gegenwärtig wenige Daten vorliegen und in denen das geschätzte Risiko, bedingt durch die Eignung des Klimas für die Übertragung von Parasiten, hoch ist. In Südamerika konnte seit 2005 ein enormer Rückgang des Erkrankungsrisikos an einer durch Bodenkontakt übertragenen Wurmerkrankung beobachtet werden. Die an die Bevölkerung angepasste Prävalenz für Askariasis, Trichuriasis und die Hakenwurminfektion liegen jeweils bei 15.6%, 12.5% beziehungsweise 11.9%.

In Kapitel 5 wird eine Weiterentwicklung der peNMIG Bayes'schen geostatistischen Variablenauswahl vorgestellt, welche die Auswahl von maximal einem Prädiktor aus einer Gruppe stark korrelierter Kovariaten erlaubt. Anhand dieses Ansatzes wurden die wichtigsten Risikofaktoren identifiziert, welche in einem Bayes'schen geostatistischen Model für die drei am stärksten verbreiteten durch Bodenkontakt übertragenen Wurmerkrankungen in Bolivien integriert wurden. Aus den Risikoschätzungen mit hoher räumlicher Auflösung geht hervor, dass 48.4% der Bolivianischen Bevölkerung mit irgendeinem durch Bodenkontakt übertragbaren Wurm infiziert ist und 2.9 Millionen annualisierte Behandlungen für eine präventive Chemotherapie nötig wären. Der Mangel an Daten legt jedoch den Bedarf einer nationalen Erhebung nahe, um eine präzisere Kartierung für die räumliche Bestimmung von Kontrollinterventionen zu ermöglichen.

Kapitel 6 beschäftigt sich mit der Analyse der raum-zeitlichen Verteilung der drei dominantesten durch Bodenkontakt übertragenen Wurmerkrankungen in Brasilien, basierend auf historischen Daten ab 1995. Unsere raum-zeitlichen Modelle beinhalten einen raum-zeitlichen Interaktionseffekt, wodurch die räumliche Struktur zeitlich variieren kann. Die Modelle basieren auf Umwelt- und sozioökonomischen Prädiktoren, welche durch eine Bayes'sche geostatistische Variablenauswahl selektiert wurden. Jenes Auswahlverfahren eignet sich für große Datensätze und berücksichtigt die Nicht-Linearität und Korrelation der erklärenden Variablen. Die raum-zeitlichen Modelle wurden innerhalb des Kontextes der integrated nested Laplace approximation/stochastisch partielle Differentialgleichungen angepasst. Wir zeigen, dass das Risiko für *A. lumbricoides* und Hakenwurminfektionen in Brasilien in den vergangenen 20 Jahren zurückgegangen ist. Das Risiko seit 2010 wurde geschätzt auf 3.6% für *A. lumbricoides*, 1.7% für den Hakenwurm und 1.4% für *T. trichiuria*. Daraus ergibt sich die Anzahl der annualisierten Behandlungen in Höhe von 1.8 Millionen, welche für eine Massenentwurmung der Kinder im Schulalter auf Gemeindeebene benötigt wird.

Diese vorliegende Thesis bietet Bayes'sche statistische Methodik für räumliche Modellierung und Schätzung des Wurmerkrankungsrisikos. Wir präsentieren und implementieren ein Bayes'sches geostatistisches Variablenauswahlverfahren, welches die Nicht-Linearität und Korrelation der Einflussvariablen berücksichtigt. Darüber hinaus werden Ansätze für raum-zeitliche und große Datensätze behandelt. Unsere wichtigsten Beiträge im Bereich der Wurm-Epidemiologie umfassen die Bereitstellung von Informationen für nationale Kontrollprogramme, um durch Bodenkontakt übertragene Wurminfektionen in Südamerika und Schistosomiasis in der Elfenbeinküste zu bekämpfen. Diese Informationen umfassen sowohl räumlich hochaufgelöste Risikoschätzungen als auch die Anzahl der Menschen, welche eine präventive Chemotherapie benötigen. Folglich bietet unsere Arbeit einen wichtigen Richtwert, von welchem weitere Schätzungen abgeleitet werden können, sobald neue Datensätze zur Verfügung stehen und Kontrollinterventionen voranschreiten.

Translated from english by Verena Jürgens

Résumé

Au cours des 20 dernières années, des efforts considérables ont été effectués afin d'amener sur le devant de la scène le combat contre les maladies tropicales négligées, dont font partie les helminthiases. Le fardeau mondial de la schistosomiase et des infections à géohelminthes dues à *Ascaris lumbricoides*, *Trichuris trichiura* et aux ankylostomes est estimé à 8.4 millions d'années de vie ajustées à l'incapacité (DALYs). Une conscience politique s'est élevée et des mesures spécifiques ont été établies afin de réduire durablement et de manière rentable ce fléau. Les moyens financiers débloqués par des philanthropes ont permis de faire avancer la recherche scientifique et les donations de vermifuges offrent une réelle opportunité d'atteindre une diminution significative du préjudice. Une connaissance et une compréhension approfondies de la distribution géographique des infections sont cruciales afin de guider et mettre en oeuvre les programmes de lutte. Des estimations à haute résolution du risque d'infections sont non seulement essentielles pour une planification rationnelle et rentable des interventions, mais aussi au cours des étapes d'évaluation et de surveillance. Des modèles statistiques peuvent être utilisés afin d'établir une relation mathématique entre les observations et les facteurs de risque, ainsi que pour prédire le risque aux endroits où l'information fait défaut. La prolifération des helminthes est régie par des interactions complexes entre des facteurs environnementaux et socio-économiques inégalement distribués dans l'espace. De plus, la présence d'un parasite à un endroit donné influence son existence dans les environs. Ainsi, les observations sont corrélées dans l'espace et la modélisation du risque doit être entreprise dans le cadre d'une approche statistique spatiale. La géostatistique Bayésienne offre une méthodologie flexible et rigoureuse pour traiter de telles données, en permettant aux modèles de prendre en compte l'incertitude spatiale.

Le premier chapitre présente quelques principes fondamentaux liés à cette thèse de doctorat portant sur les "Prédictions fondées sur un modèle géostatistique des risques d'helminthiases pour assister les interventions de contrôle". Un aperçu de la biologie et

de l'épidémiologie des schistosomes et des géohelminthes étudiés est offert et les questions liées aux traitements et aux outils de diagnostic actuellement disponibles sont évoquées. Les stratégies de lutte sont abordées, de même que les bases de la méthodologie géostatistique Bayésienne pour la modélisation du risque d'infections. Le but général de ce travail est de développer, mettre en œuvre et valider la méthodologie géostatistique Bayésienne dans un cadre épidémiologique. L'accent est mis sur l'estimation de la répartition géographique du risque d'helminthiases avec comme but ultime de combler les lacunes à la fois méthodologiques et épidémiologiques empêchant de répondre aux besoins urgents de cibler spatialement les interventions.

Les problématiques importantes concernant la modélisation spatiale du risque d'infections, et plus particulièrement les difficultés computationnelles liées à l'analyse Bayésienne de larges jeux de données, l'hétérogénéité des données historiques, les hypothèses de stationnarité et d'isotropie, ou encore les méthodes de sélection de variables explicatives sont passés en revue dans le Chapitre 2. Une emphase particulière est mise sur la sélection de variables dans un contexte géostatistique Bayésien avec une paramétrisation des coefficients de régression sous forme d'un mélange de distributions normale et inverse-gamma (peNMIG). Cette approche est appliquée à des données historiques relatives aux infections à *Schistosoma mansoni* en Côte d'Ivoire et est comparée à des méthodes de sélection plus traditionnelles. Cette novatrice paramétrisation permet une sélection rigoureuse des variables catégoriques, identifie les prédicteurs importants du risque d'infection dû à *S. mansoni* et conduit à un modèle plus parcimonieux comparé aux approches traditionnelles de sélection qui ignorent la structure spatiale des données.

Le Chapitre 3 propose une analyse du statut des infections liées à *S. mansoni* et *S. haematobium* qui ont été collectés lors d'une étude transversale réalisée entre novembre 2011 et février 2012 sur les enfants d'âge scolaire en Côte d'Ivoire. Un modèle Bayésien géostatistique multinomial a été appliqué pour estimer chaque risque de mono-infection, ainsi que la co-infection. Le risque général des schistosomiases est estimé à 8.9% (5.3% pour *S. haematobium* et 3.8% pour *S. mansoni*) et 2 millions de traitements seraient nécessaires pour une intervention médicamenteuse préventive au niveau des districts sanitaires. Les répartitions spatiales distinctes de *S. haematobium* et *S. mansoni* impliquent que la co-infection par les deux types de parasites est faible à travers le pays.

Le Chapitre 4 présente les résultats d'une revue systématique et d'une méta-analyse géostatistique des données de prévalence des géohelminthes pour toute l'Amérique du Sud. Sur 4 085 articles scientifiques examinés, 174 ont fourni des données pertinentes. Très peu

d'observations ont pu être réunies pour le sud et la côte ouest du sous-continent. Aucune donnée n'a été identifiée pour l'Uruguay et seul un nombre limité d'information a été récolté pour le Suriname, le Guyana, la Guyane et l'Equateur. Au total, des prévalences d'ascaridiose, de trichurose et d'ankylostomose ont été extraites pour 6 948 localités. Les larges sets de données ont conduit à développer des modèles Bayésiens géostatistiques où l'incertitude spatiale est estimée par la méthode du "processus de prévision". L'estimation des risques pour le sous-continent sud-américain offre un important support pour le ciblage spatial des interventions de lutte contre les géohelminthiases et suggère le besoin de collecter des informations supplémentaires, en particulier dans les pays où les données disponibles sont rares et où le risque estimé, motivé par des conditions climatiques favorables, est élevé. En Amérique du Sud, le risque d'être infecté par un géohelminthe a considérablement diminué depuis 2005 et la prévalence (ajustée à la population) est estimée à 15.6%, 12.5% et 11.9% pour l'ascaridiose, la trichuroses et l'ankylostomose, respectivement.

Dans le Chapitre 5, la sélection de variable Bayésienne géostatistique peNMIG a été d'avantage développé afin de permettre la sélection d'un seul prédicteur au maximum parmi des groupes de variables hautement corrélées. Cette approche a permis d'identifier les facteurs de risques importants et a abouti à la construction de modèles Bayésiens géostatistiques du risque des trois principales géohelminthiases en Bolivie. Les estimations des risques à haute résolution spatiale indiquent que 48.4% de la population bolivienne est porteuse d'une géohelminthiase et que 2.9 millions de traitements par année seraient requis pour une intervention chimiothérapique. Cependant, le faible nombre d'études disponibles en Bolivie suggère le besoin d'une enquête nationale afin d'obtenir une cartographie du risque qui serait à même de cibler plus précisément les interventions.

A partir de données historiques antérieures à 1995, la distribution spatio-temporelle des trois principales géohelminthiases au Brésil est analysée dans le Chapitre 6. Nos modèles spatio-temporaux considèrent une interaction espace-temps qui permet à la structure spatiale de varier au cours du temps. Les modèles sont construits sur des prédicteurs environnementaux et socio-économiques sélectionnés par une sélection de variables géostatistique Bayésienne pour de larges données, qui adresse la non-linéarité et la corrélation des variables explicatives. Les modèles spatio-temporaux ont été estimés à l'aide des approximations de Laplace et des équations aux dérivées partielles stochastiques. Nous montrons que les risques d'ascaridiose et d'ankylostomose ont diminué au cours des 20 dernières années au Brésil. A partir de 2010, le risque est estimé à 3.6% pour *A. lumbricoides*, 1.7% pour les ankylostomes, et 1.4% pour *T. trichiura*. Ainsi, le nombre de traitements annuels

nécessaire à un déparasitage de masse des enfants d'âge scolaire au niveau des municipalités est estimé à 1.8 millions.

Le travail présenté dans cette thèse propose une méthodologie statistique Bayésienne pour la modélisation spatiale et l'estimation des risques d'helminthiases. Nous proposons et implémentons une sélection de variables géostatistique Bayésienne qui permet la sélection de prédicteurs non-linéaires et corrélés, ainsi que des approches de modélisations spatio-temporelles et de grands ensembles de données. Notre principale contribution dans le domaine de l'épidémiologie des helminthes est d'informer les programmes nationaux de contrôle avec des estimations du risque et du nombre de personnes ayant besoin d'une chimiothérapie préventive pour lutter contre les géohelminthiases en Amérique du Sud et les schistosomiases en Côte d'Ivoire. Ainsi, notre travail fournit une référence importante sur laquelle d'autres estimations pourraient être dérivées, au fur et à mesure que de nouvelles données seront disponibles et que les interventions de contrôle progressent.

Acknowledgements

We are wont to say that time flies when we enjoy it. I remember the day I have knocked on the door of the Swiss Tropical and Public Health institute (Swiss TPH) as if it were yesterday and I had the greatest pleasure to work during these past three years with a number of people I would like to acknowledge

I would like to express my warmest thanks and my profound gratitude to my supervisor, Dr. Penelope Vounatsou. She took me into her Bayesian world and guided me all the way along. It has been a great chance to work at her side and benefit from her statistical expertise. Her dedication and commitment to work helped me to push my own limits and I am deeply grateful for the confidence she has placed in me and for the precious time that she has given me. I would also like to sincerely thank Prof. Dr. Jürg Utzinger for his co-supervision, his enthusiasm, his encouragements, and his valuable comments on my work. It has been a most rewarding experience to learn from his epidemiological expertise. Jürg was also kind to magnify all my manuscripts with his magic touch. The success of my PhD thesis owes a lot to this synergic team. Thank you very much to both of you, Penelope and Jürg, for teaching me what I learned. I could not have thought of better guides!

I am grateful to Prof. Dr. Marcel Tanner, director of the Swiss TPH, for leading a stimulating environment and for making me feel proud to be part of it. I would like to extend my thanks to Dr. Robert Bergquist who kindly agreed to evaluate my work as an external reviewer and to Prof. Dr. Reto Brun who accepted to chair my PhD defence.

It is my pleasure to acknowledge my current and previous colleagues at Swiss TPH for their support and the good atmosphere in the office: Abbas Adigun, Amek Ombek, Christian Hermann, Eric Diboulo, Erika Muller, Patricia Biedermann, Ronaldo Scholte, Sabelo Dlamini, Sammy Khagayi, Simon Kasasa, Susan Rumisha, and Yingsi Lai. My special thanks are addressed to Nadine Schur, Laura and Dominic Gosoniu, Federica Giardina, and Verena Jürgens for their previous work that made me grow up faster and to Alex

Karagiannis for his help and our various knowledge exchanges. I would also like to thank Eveline Hürlimann for her invaluable contribution to this unforgettable journey.

I am also indebted to my Ivorian partners from the Centre Suisse de Recherches Scientifiques (CSRS) for their warm support during my visit in Abidjan. I would especially like to thank Dr. Giovanna Raso for her contribution with schistosomiasis data analysed in this work. The CSRS is a big family and I cannot mention everybody here. However, it would be unfair not to mention Prof. Dr. Bonfoh, Dr. Mathurin Koffi, Dr. Silué Kigbafori, Dr. Jean Coulibaly, Clarisse Houngbedgy, Richard Yapi, Daniele Konan, and Christelle Dassi for their help in data collection, data geolocalisation and field understanding.

I am also keen to acknowledge the anonymous reviewers of Acta Tropica, The Lancet Infectious Diseases, Parasites & Vectors and PLoS Neglected Tropical Diseases for their constructive comments on my manuscripts. Additional thanks are addressed to The Lancet Infectious Diseases and Brett Ryder from The Heartagency for their kind authorization to use the image that illustrates the cover of this thesis.

Finally, I do not forget to thank from the bottom of my heart my partner, parents, family and friends who have always encouraged me in my choices.

This work was financed by a PhD scholarship from the CSRS foundation. The "Reisefonds für den akademischen Nachwuchs" from University of Basel covered the cost linked to the presentation of part of my work at the Spatial Statistics Conference 2013 in Columbus, Ohio, USA and the "Stiftungsrat" of the "Basler Studienstiftung" has financed the printing of this thesis. I am very grateful to all for their generous funding.

Contents

List of Figures

List of Tables

List of Abbreviations

ACT	Artemisinin-based Combination Therapy
BCI	Bayesian Credible Interval
CI	Confidence Interval
CSRS	Centre Suisse de Recherches Scientifiques
DALY	Disability-Adjusted Life Year
DHS	Demographic and Health Survey
DIC	Deviance Information Criterion
EVI	Enhanced Vegetation Index
GAHI	Global Atlas of Helminth Infections
GDB	Global Disease Burden
GIS	Geographical Information System
GMRF	Gaussian Markov Random Field
GNTD	Global Neglected Tropical Diseases (database)
GPS	Global Positioning System
HDI	Human Development Index
HII	Human Influence Index
IMR	Infant Mortality Rate
ICOSA	Integrated Control of Schistosomiasis in Sub-Saharan Africa
IEC	Information, Education and Communication
INLA	Integrated Nested Laplace Approximation
LST	Land Surface Temperature
MAE	Mean Absolute Error
MCMC	Markov Chain Monte Carlo
MDG	Millennium Development Goals
ME	Mean Error
MICS	Multiple Indicator Cluster Surveys

MoH	Ministry of Health
NDVI	Normalised Difference Vegetation Index
NTD	Neglected Tropical Disease
OR	Odds Ratio
PAHO	Pan American Health Organization
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
peNMIG	parameter expanded Normal Mixture of Inverse-Gamma
RS	Remote Sensing
SCI	Schistosomiasis Control Initiative
SDG	Sustainable Development Goals
SCORE	Schistosomiasis Consortium for Operational Research and Evaluation
SD	Standard Deviation
SPDE	Stochastic Partial Differential Equation
STH	Soil-Transmitted Helminth
Swiss TPH	Swiss Tropical and Public Health (institute)
UBN	Unsatisfactory Basic Needs
WHO	World Healh Organization
WHS	World Health Surveys

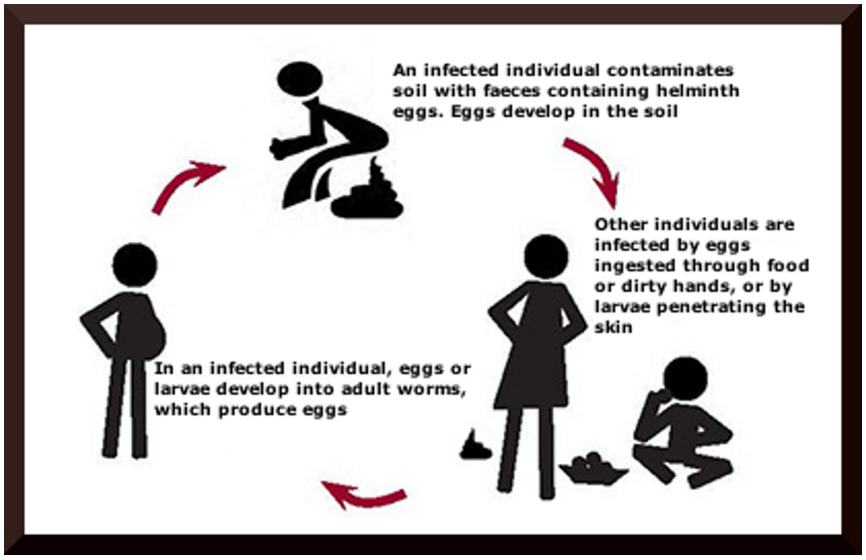
Chapter 1

Introduction

The purpose of this introduction is to lead the reader into the topic of this PhD thesis focussed on "Geostatistical model-based predictions of helminthiases risk to assist control interventions". Section 1.1 and 1.2 introduce the biology and epidemiology of schistosomes and soil-transmitted helminths, as well as available diagnoses and treatments. Public health control measures established to fight the burden are presented in Section 1.3, while Section 1.4 addresses the basics of the statistical methodology involved in spatial modelling of the geographical distribution and risk mapping.

1.1 Soil-transmitted helminth infections

Soil-transmitted helminth (STH) infections are named after their mode of transmission through faeces-contaminated soil and are part of the 17 diseases listed as "Neglected Tropical Diseases" (NTD) by the World Health Organization (WHO). The main species affecting human are the roundworm *Ascaris lumbricoides*, the whipworm *Trichuris trichiura* and the two hookworm species *Ancylostoma duodenale* and *Necator americanus*.



Source: WHO (2011)

Figure 1.1: Soil-transmitted helminths schematic life cycle.

1.1.1 Life cycles and morphological characteristics

Soil-transmitted helminths infect human intestines. Each day they produce thousands of eggs which are released in the soil through the faeces. If environmental conditions are

suitable, eggs further develop into infective stages and are ready to infect another human host (Figure 1.1). *A. lumbricoides* and *T.trichiura* present similar transmission modes. Eggs are ingested *via* soil-contaminated hands or food, larvae hatch in the human host intestine, and adult worms release eggs in the soil through the faeces to complete the cycle. For hookworm, larvae hatch in the soil and penetrate the body through the skin. They reach a mature stage inside the host intestine and release eggs which are passed in the stool.

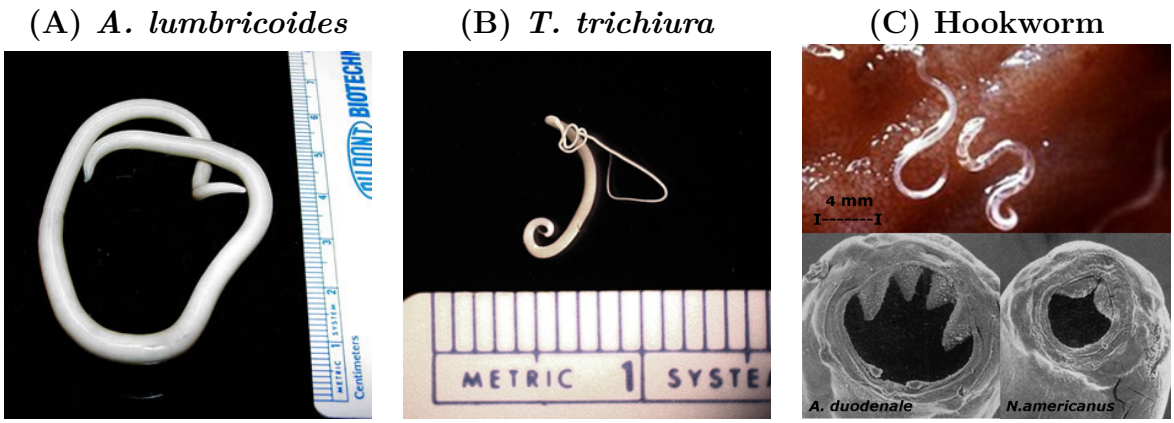
Soil-transmitted helminth eggs measure a few dozen micrometers and present different morphological characteristics as depicted in Figure 1.2. Fertilised *A. lumbricoides* eggs have a mammalated thick outer shell, *T. trichiura* eggs exhibit a distinctive "lemon" shape with a plug at each pole, and hookworm eggs show a thin shell and visible embryonic cleavages. Visual examination of hookworm eggs does not allow differentiating between the *A. duodenale* and *N. americanus* species.

Adult worms differ in size and shape (Figure 1.3). The *A. lumbricoides* worm is creamy white/pinkish and the female can measure up to 40 cm. *T. trichiura* is smaller and exhibits a characteristic whip-shape. Hookworms are bent as a hook and present a head with a well developed mouth, which differenciates the two species *A. duodenale* and *N. americanus*.

(A) *A. lumbricoides* (fertilised) **(B) *T. trichiura*** **(C) Hookworm**

Source: http://www.dpd.cdc.gov/dpdx/HTML/Image_Library.htm; accessed: December 2013

Figure 1.2: *A. lumbricoides* (A), *T. trichiura* (B), and hookworm (C) eggs in an unstained wet mount of stool.



Sources: http://www.dpd.cdc.gov/dpdx/HTML/Image_Library.htm (A, B and C top); accessed: December 2013 and Hotez (1995) (C bottom)

Figure 1.3: *A. lumbricoides* (A), *T. trichiura* (B), and hookworm (C) adult worms.

1.1.2 Clinical conditions

Soil-transmitted helminth infections are often a chronic condition and may be asymptomatic. Abdominal discomfort (diarrhoea, pain) and general weakness may occur, depending on the intensity of infection. The most serious complications are rectal prolapse and intestinal obstruction due to *A. lumbricoides* and *T. trichiura* heavy infections. Attachment of adult hookworms to the intestine causes loss of blood, which can lead to serious consequences in developing anaemia and protein deficiency. Repeated exposures, chronic conditions and high parasite load of school-aged children are worrisome. The negative impact of soil-transmitted helminthiases on nutritional status, cognitive and physical development, and school performances has consequences for a lifetime. Moreover, as parasites affect primarily deprived people living in poor socio-economic conditions, infection can be seen as a cause and a consequence of poverty, locking vulnerable people in a vicious circle.

1.1.3 Diagnosis

As soil-transmitted helminth infections lead to non-specific symptoms, passive case detection is not reliable. Active detection is performed by coprological examination within the surveyed communities. Eggs can be detected in the faeces using formalin-ether concentration or sedimentation techniques. The most widely used diagnostic tool for detecting soil-transmitted helminth infections is the WHO recommended Kato-Katz technique. A standardised thick smear is prepared and experienced technicians identify the different stained eggs under a microscope. The Kato-Katz owes its popularity to its easy implementation under field conditions at a relatively low cost, estimated to 0.04 US$ for a duplicate

Kato-Katz (salaries and infrastructure expenses are excluded from this calculation) (Speich et al., 2010). Furthermore, it allows quantifying intensity by counting the number of eggs per gram of faeces. However, the method is largely criticised and presents severe drawbacks; the most important being the lack of sensitivity, which can be increased by multiplying the thick smears from the same stool sample, or even better, by examining stool sample collected over consecutive days (Knopp et al., 2008). In addition, slides, which degrade fast (around 30 minutes after preparation), should be read immediately to be able to detect hookworm eggs.

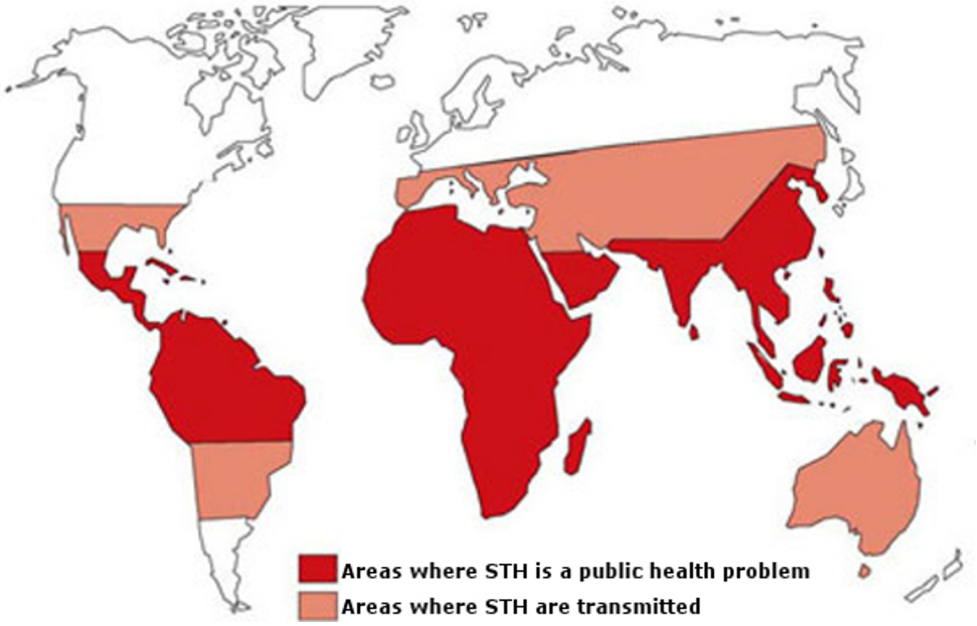
Since the sensitivity of the diagnostic tests is positively correlated to intensity and in view of an expansion of low intensity status due to progresses of control measures, alternative diagnostic techniques are needed to match better sensitivity, low cost and feasibility in the field. Flotation techniques, such as the Mc Master and its derivatives FLOTAC and mini-FLOTAC, are currently investigated. The FLOTAC techniques show a better sensitivity compared to triplicate Kato-Katz (Knopp et al., 2009). Additionally, samples can be stored for later examination and faecal material is manipulated in a closed environment, which increases safety and comfort of the technicians. Substitution of traditional microscopy by a mobile phone with an additional lens is being explored and might be attractive in the future (Bogoch et al., 2013).

1.1.4 Treatment

WHO recommends four anthelmintic drugs for large-scale chemotherapy against soil-transmitted helminthiases: albendazole, mebendazole, lavamisole and pyrantel pamoate. However, only the first three are recommended for school deworming programmes because of their facility of administration as a single dose, irrespective of the child's weight. These drugs are inexpensive (0.01–0.02 US$) and safe (WHO, 2006). Indeed, they benefit from long term implementation, have been largely tested and show few minor side-effects. However, the risk with mass drug administration is the development of resistance. Therefore, drug efficacy should be closely monitored (WHO, 2013a). Keiser and Utzinger (2008) assessed the efficacy of single-dose administration of albendazole, mebendazole, levamisole and pyrantel pamoate against soil-transmitted helminthiases in a meta-analysis of 20 randomised controlled trials. Ascariasis was effectively treated with all drugs, albendazole showed more efficacy than the three other drugs against hookworm, and low cure rate was generally observed for *T. trichiura*.

Research for novel anthelmintic drugs is on-going and next-generation compounds are

eagerly awaited (Keiser and Utzinger, 2010). Despite that, clinical trials with tribendimidine have demonstrated a broad anti-parasitic spectrum, efficacy against trichuriasis and hookworm remains unsatisfactory (Xiao et al., 2013). In parallel, the development of preventive vaccines against hookworm progresses. Two candidate antigens stimulating the human immune system to produce antibodies that inhibit hookworm blood feeding are in Phase I clinical trial (Beaumier et al., 2012). The approach of the future seems to be heading towards a synergic strategy vaccine–chemotherapy (Loukas et al., 2006).



Source: WHO (2002a)

Figure 1.4: Global distribution of soil-transmitted helminth infections.

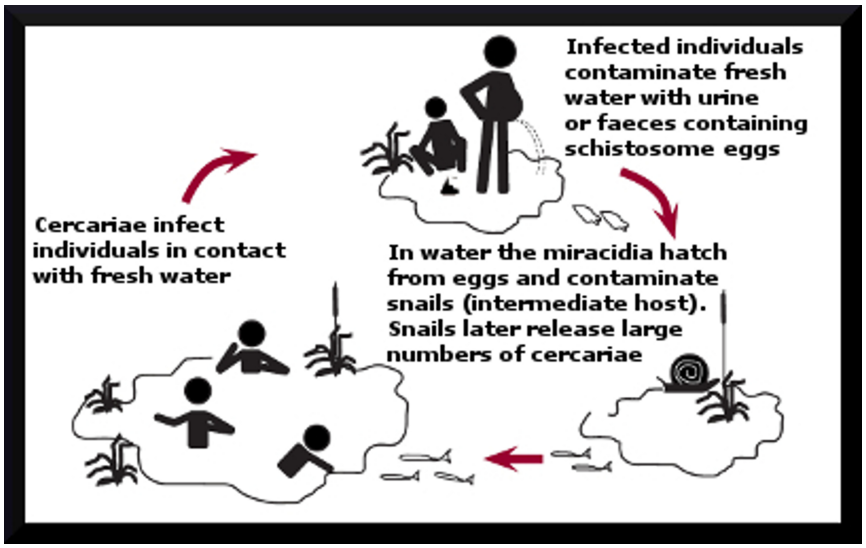
1.1.5 Global distribution and disease burden

Soil-transmitted helminths are widely distributed in tropical and sub-tropical areas of sub-Saharan Africa, Asia, and Latin America (Figure 1.4). The latest update of the global picture estimated the number of people infected to 644 million (30%), 457 million (21%), and 355 million (16%) for *A. lumbricoides*, hookworm and *T. trichiura*, respectively (De Silva et al., 2003). To quantify the health-impact of this dramatic figure, indicators such as mortality and disability-adjusted life year (DALY) — which expresses the number of years lost due to ill-health, disability or early death — have to be considered. In a single

year, hookworm, *A. lumbricoides*, and *T. trichiura* infections are responsible for the death of 65,000, 60,000, and 10,000 people, respectively (WHO, 2002b). A fatal outcome remains rare, given the prevalence of the infections. The 2010 global disease burden (GDB) study estimated that hookworm, *A. lumbricoides*, and *T. trichiura* cause 3.2, 1.3, and 0.6 million DALYs (Murray et al., 2013). Thus, soil-tranmitted helminthiases carry the largest DALY among the neglected tropical diseases and contribute to 20% of their global burden.

1.2 Schistosomiasis

Human schistosomiases, also known as bilharzias, are blood-flukes infections due to different trematode species of the genus *Schisotosma*. They can be divided into intestinal schistosomiasis (*S. mansoni* and *S. intercalatum*), urinary schistosomiasis (*S. haematobium*), and Asian intestinal schistosomiasis (*S. japonicum* and *S. mekongi*).



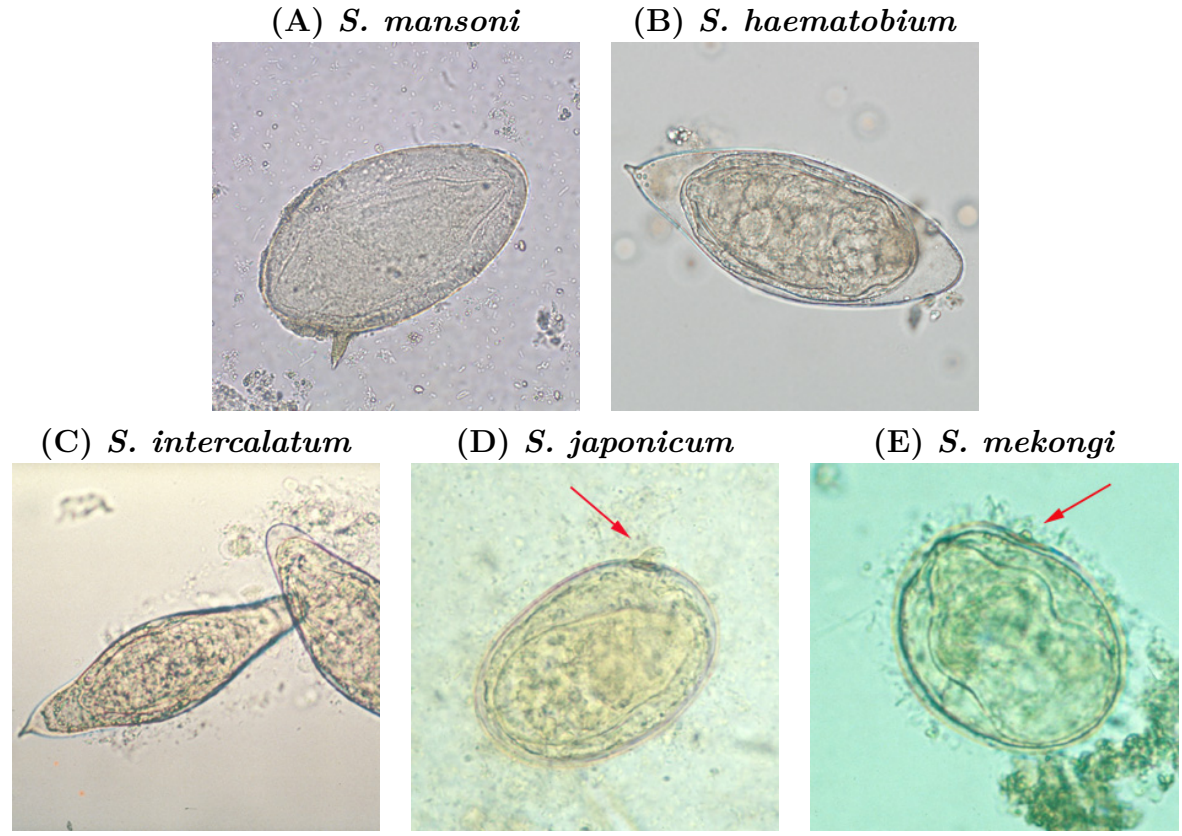
Source: WHO (2011a)

Figure 1.5: *Schistosoma* schematic life cycle.

1.2.1 Life cycles and morphological characteristics

Cercariae penetrate human skin and schistosomula forms migrate to the liver *via* the blood circulation where they reproduce and mature. Adult worms reach the mesenteric vessels of bowel (*S. mansoni*, *S. japonicum*, *S. intercalatum*, and *S. mekongi*) or bladder (*S. haematobium*) and females release eggs, which either are trapped in the tissues or

excreted in faeces or urine. Eggs hatch in contact with water into miracidia and further infect snails of the genus *Biomphilaria* or *Bulinus*. After maturation, cercarial larvae are liberated into fresh water and complete the life cycle (Figure 1.5). Schistosome eggs are depicted in Figure 1.6. *S. mansoni* eggs (shed in stool) present a characteristic prominent lateral spine, while *S. haematobium* eggs (shed in urine) show an apparent terminal spine. *S. intercalatum* eggs are rather similar to *S. haematobium*, but they are restricted to East-Central Africa and found in stool. *S. japonicum* and *S mekongi* eggs are rounded and present a less conspicuous spine; these species are found in the Far East, *S. mekongi* being restricted to the area along the Mekong River.



Source: http://www.dpd.cdc.gov/dpdx/HTML/Image_Library.htm; accessed: December 2013

Figure 1.6: *S. mansoni* (A), *S. haematobium* (B), *S. intercalatum* (C), *S. japonicum* (D), and *S. mekongi* (E) eggs in an unstained wet mount.

1.2.2 Clinical conditions

Earliest symptoms, which may often go unnoticed, consist of a skin rash or a prurit. A few weeks later, flu-like manifestations may appear such as fever, chills, cough and muscle pain. In the absence of treatment, worms can persist for years and release a massive quantity of eggs, which in turn will damage different tissues. Urinary schistosomiasis affects the perivesical tissues, causing red-blood cell loss in urine (haematuria), trouble passing urine (dysuria) and more severe consequences such as bladder calcification, genital tract lesions, nephrosis, or bladder cancer. Intestinal schistosomiasis alters peri-intestinal tissue and creates abdominal pain, intestinal bleeding and diarrhoea, accumulation of fluid in the peritoneal cavity, splenomegaly and hepatomegaly. Rarely, eggs reach the brain or the spinal medulla and induce seizures, paralysis, or spinal cord inflammation. All chronic schistosomiasis infections are also responsible for growth retardation, impairment of cognitive development and contribute to anaemia.

1.2.3 Diagnosis

Schistosome eggs can be detected in stool (for *S. mansoni*, *S. intercalatum*, *S. japonicum* and *S. mekongi*) or in urine (for *S. haematobium*). Stained eggs are quantified by microscopy on Kato-Katz slides or after filtration in case of urine testing. *S. haematobium* can also be easily detected with semi-quantitative reagent strip indicating the presence of blood and proteins in urine, while self-reported blood questionnaires are a useful tool for a rapid assessment in high risk communities. However, sensitivity and specificity drawbacks of the aforementioned tests are well established due to daily variations in egg excretion (optimal at noon), heterogeneous eggs distribution within the stool, or shortcomings inherent to the tests per se (Utzinger et al., 2001; Booth et al., 2003; Enk et al., 2008).

Issues regarding the reliability to detect the infection in low-endemic settings have led to the development of antigen-based tools. For example, circulating cathodic antigen tests, which detect the presence of antibodies against *S. mansoni* and *S. haematobium* in a finger-prick blood or urine sample, have been tested and showed good sensitivity (Coulibaly et al., 2011).

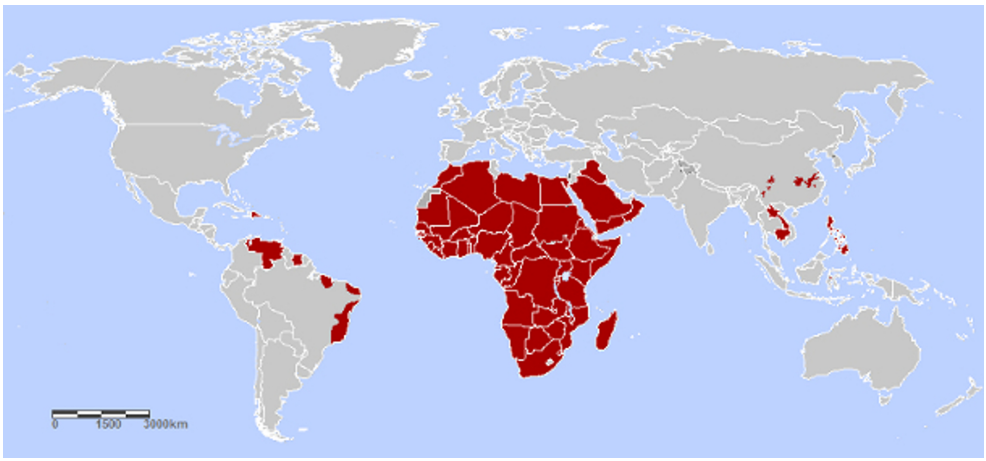
1.2.4 Treatment

Praziquantel (40 mg/kg) is currently the treatment of choice against schistosomiasis. This molecular compound acts primarily on the adult worms and is generally well tolerated despite frequent abdominal side effects. However, it should not be prescribed to patients

with epilepsy or central nervous system involvement within a mass distribution. Required dose of praziquantel can be extrapolated from patient height with a pole-dose developed by WHO (Montresor et al., 2005). The cost is approximately 0.20 US$ per dose (WHO, 2006). Two doses administrated a few weeks apart showed a higher cure rate, especially regarding *S. haematobium* (Garba et al., 2012). Oxamniquine is another drug with an activity on *S. mansoni* and derivatives of artemisinin, primarily used as antimalarials, have shown some potential against schistosomiasis (Utzinger et al., 2010). Candidate vaccines are currently tested in clinical trials (Hotez, 2011). Hopefully, a preventive vaccine which satisfies safety and efficacy exigencies will be released in the future.

1.2.5 Global distribution and disease burden

Schistosomiasis is endemic in 76 tropical and sub-tropical countries of Africa, Asia and Latin America (Figure 1.7). In 2003, population at risk and number of people affected by schistosomiasis have been estimated to 779 million and 207 million, respectively (Steinmann et al., 2006). In sub-Saharan Africa alone, 280,000 deaths per year are attributed to kidney failure or haematemesis caused by schistosomiasis (Van der Werf et al., 2003). The 2010 GDB study estimated that schistosomiasis is responsible for a burden of 3.3 million DALYs (Murray et al., 2013), which represents 13% of the global NTDs burden. Thus, schistosomiasis is ranked second in terms of DALY importance among the NTDs (*ex aequo* with leishmaniasis and after soil-transmitted helminth infections).



Source: http://gamapserver.who.int/mapLibrary/app/searchResults.aspx; accessed: December 2013)

Figure 1.7: Areas endemic for schistosomiasis.

1.3 Helminthiases control strategies

The ultimate goal of the fight against helminthiases is to eradicate the infections and offer the next generations the prospect of a world free of worms. Reduction of their burden is mainstreamed into the Millennium Development Goals (MDGs) established to help least-developed countries to lift out of poverty. However, the magnitude of the task, limited financial resources, and the heterogeneous spatial distribution of the infections temporarily forced some control programmes to revise their objectives. Thus, a two-tier approach has been developed targeting i) a control of morbidity prioritising a regular chemotherapy administration to school-aged children in high endemic areas and ii) a control of transmission where risk is lower (Engels et al., 2002).

Control strategies advocated by WHO include three measures which are i) repeated mass drug administration, ii) sanitation improvement, and iii) health education (WHO, 2010). For schistosomiasis, an additional intervention consists of controlling the snails in order to interrupt transmission cycle. Although Japan and the People's Republic of China were successful while acting on snails with molluscicides (Zhang and Jiang, 2011), snail population control is costly, difficult to implement and pollutes with serious consequences to the ecosystem (Fenwick and Savioli, 2011). A malacological approach is thus confined to restricted actions in space for improving control at local level (Rollinson, 2009).

Mass periodic chemotherapy aims to reduce intensity of infections and decrease the amount of eggs released in the environment. School-aged children are the prioritised target of preventive chemotherapy. First, because a child needs more than any other essential nutrients to grow and learn, and second, because they are highly exposed while playing with soil and water. Preventive chemotherapy guidelines have been established for different risk zones (see Table 1.1 for soil-transmitted helminthiases and Table 1.2 for schistosomiasis).

Many schistosomiasis national control programmes have yielded encouraging results and several have proved that elimination is an achievable goal (Rollinson et al., 2013). However, some countries, mainly in Sub-Saharan Africa, still have a long way to reach elimination and are at an early stage of the morbidity control process. Regarding soil-transmitted helminths, existing drugs are known to lack efficacy and re-infection is high (Keiser and Utzinger, 2008). While waiting for the discovery of new chemical components, mass drug therapies with existing treatments, together with improved access to clean water, adequate sanitation and health education are implemented to lower parasite load and enable morbidity control.

One of the most important aspects of every control initiative is to make its programme

Table 1.1: WHO-recommended treatment strategy for soil-transmitted helminths in preventive chemotherapy.

Category	**Prevalence of any STH infection among school-aged children**	**Action to be taken**	
High-risk community	$\geq 50\%$	Treat all school-aged children (enrolled and not enrolled) twice[1] each year	Also treat: • preschool children; • women of childbearing age, including pregnant women in the $2nd$ and $3rd$ trimesters and lactating women; • adults at high risk in certain occupations (e. g. tea-pickers and miners)
Low-risk community	$\geq 20\%$ and $< 50\%$	Treat all school-aged children (enrolled and not enrolled) once each year	Also treat: • preschool children; • women of childbearing age, including pregnant women in the $2nd$ and $3rd$ trimesters and lactating women; • adults at high risk in certain occupations (e. g. tea-pickers and miners)

[1] If resources are available, a third drug distribution intervention might be added. In this case the appropriate frequency of treatment would be every 4 months.

General remark: When prevalence of any STH infection is less than 20%, large-scale preventive chemotherapy interventions are not recommended. Affected individuals should be dealt with on a case-by-case basis.

Source: WHO (2011a).

Table 1.2: WHO-recommended treatment strategy for schistosomiasis in preventive chemotherapy.

Category	**Prevalence among school-aged children**	**Action to be taken**	
High-risk community	$\geq 50\%$ by parasitological methods (intestinal and urinary schistosomiasis) or $\geq 30\%$ by questionnaire for visible haematuria (urinary schistosomiasis)	Treat all school-aged children (enrolled and not enrolled) once a year	Also treat adults considered to be at risk[1]
Moderate-risk community	$\geq 10\%$ but $< 50\%$ by parasitological methods (intestinal and urinary schistosomiasis) or $< 30\%$ by questionnaire for visible haematuria (urinary schistosomiasis)	Treat all school-aged children (enrolled and not enrolled) once every 2 years	Also treat adults considered to be at risk[2]
Low-risk community	$< 10\%$ by parasitological methods (intestinal and urinary schistosomiasis)	Treat all school-aged children (enrolled and not enrolled) twice during their primary schooling age	Praziquantel should be available in dispensaries and clinics for treatment of suspected cases

[1] From special groups to entire communities living in endemic area.

[2] Special risk groups only.

Source: WHO (2011a).

sustainable and cost-effective. On one hand, disease distributions often overlap (Yajima et al., 2011), control strategies share similarities and treatments can be combined. Coordination of the different control programmes optimises the costs, while fighting poverty with a common objective of morbidity reduction due to (poly) parasitism (Molyneux et al., 2005; Hotez et al., 2007). On the other hand, the success of a control programme highly depends on the integration of treatments and strategies into all existing health and social systems. Indeed, health systems must be able to efficiently deliver the required services, while medical sectors, educational institutions and water and sanitation engineering must work in close collaboration. Thus, a control programme should not be vertically thought, but should be truly integrated and lead to a strengthened health system (Utzinger et al., 2009; Gyapong et al., 2010). In addition, careful planning, reliable analysis of the situation, constant monitoring, and evaluation of interventions are of most importance for operational efficiency and long-term elimination efforts (Yekutiel, 1981; Brooker et al., 2004a).

1.4 Statistical modelling and infection risk mapping

In-depth knowledge and understanding of the spatial distribution of the infections are crucial to guide and implement control programmes. High spatial resolution estimates of the risk of infection allow determining the number of people infected at fine-scale. The number of people infected can be subsequently aggregated to provide population-adjusted risks at any geographic area of interest. It is then straightforward to quantify the population living in high risk areas, as well as the number of treatments required for interventions.

1.4.1 Epidemiological data

Reports of spatially explicit prevalence of the infections, i. e. number of people infected out of a sample of people at given locations, are needed for building up data-driven spatial statistical models. A good coverage of the geographical area of interest is of most importance to enable models to capture heterogeneity in the risk surface and to predict the risk at unobserved locations with low uncertainty. Such point level data can be described as geostatistical and the risk is assumed to be potentially measurable all along a continuous surface.

In front of the lack of readily available georeferenced survey data pertaining to helminthiases infections, initiatives have been launched to build up databases with the Global Neglected Tropical Diseases Database (GNTD, www.gntd.org) and the Global Atlas of

Helminth Infections (GAHI, www.gahi.org). Data from the GNTD are systematically extracted from grey and published literature following a standard protocol, retrospectively georeferenced and collated into an open-access platform.

1.4.2 Risk factors

Various climatic and socioeconomic factors known to play a role in helminths transmission can be considered to explain the observed prevalence. Environmental factors related to temperature and soil characteristics are available from remote sensing (RS) that uses radiations captured by satellites to derive high spatio-temporal resolution maps. Weather stations, census and other standardised surveys provide data which can be spatially extrapolated to provide proxies related to precipitation, poverty or risk behaviours. Digitalised maps are another source of information which can be used, for example, to inform about the localisation of water bodies. Manipulation of all those data requires global positioning systems (GPS) and geographical information system (GIS) technologies.

1.4.3 Bayesian geostatistical modelling

Helminth's proliferation is governed by complex interactions of environmental and socioeconomic factors heterogeneously distributed in space. In addition, the presence of parasites at a given location influences its occurrence in the neighbourhood. Statistical models can be used to explain and predict the risk of infection. Due to the spatial structure of the data, modelling has to be handled within a spatial statistical approach.

Geostatistical models are used to capture the relationship between the risk and covariates, while accounting for spatial uncertainty (Cressie, 1990; Diggle et al., 1998). The general idea is to see an observation collected at a given location as a realisation of a random variable which, conditionally on an underlying latent spatial process, is generated by a data distribution (e. g. binomial). An appropriate transformation (logit) of the location parameter of the data distribution is then expressed as a linear function of the effects of covariates and the realisation of an unobserved latent variable. The spatial process is estimated by introducing random parameter (latent data at locations) assumed to arise from a Gaussian distribution with complex covariance matrix defined as a function of the distance between all the observed locations.

Such models are highly parameterised and model parameters cannot be estimated by maximum likelihood methods. Bayesian framework offers a flexible approach for complex models and allows the data-process-parameters joint distribution to be decomposed into

different hierarchical levels (Banerjee et al., 2003). Markov chain Monte Carlo (MCMC) algorithms estimate model parameters by drawing samples of marginal posterior distributions through iterative sampling. Starting from an arbitrary point, they construct a Markov chain which, after reaching equilibrium, provides a range of likely values which can be summarised by different measures, such as the median and the standard deviation. Common MCMC methods are the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) and its special case, the Gibbs sampler (Gelfand and Smith, 1990). MCMC sampling algorithms may converge slowly, especially for geostatistical models highly parameterised. Furthermore, handling large datasets is computationally challenging. Methods based on data dimension reduction, such as the Gaussian predictive process (Banerjee et al., 2008) and the fixed-rank kriging (Cressie and Johannesson, 2008), or based on covariance tapering (Furrer et al., 2006) have been proposed to overcome this issue. The Gaussian predictive process approximation estimates the spatial process from a subset of locations (knots). The method is popular due to its flexibility but can be sensitive to the selection and number of knots.

Another approach has recently emerged for Bayesian geostatistical inference. In brief, the latent Gaussian process modelled with a Matérn covariance function is substituted by a Gaussian Markov random field (GMRF) constructed explicitly using stochastic partial differential equation (SPDE) (Lindgren et al., 2011). The Markovian properties of the GMRF imply that most nodes only conditionally depend on their neighbours and that the precision matrix is sparse. Thus, fast numerical algorithms for sparse matrices can be used, which enable to bypass complex integrations and approximate the posterior distribution using integrated nested Laplace approximation (INLA) (Rue et al., 2009).

1.4.4 Helminthiases mapping

Bayesian geostatistics have been widely applied in disease risk mapping and efforts to produce smooth maps of helminthiases risk have ramped up in recent years. Mapping has mainly concentrated in Africa. Existing maps cover small geographical areas (Raso et al., 2005, 2006a), countries (Clements et al., 2009a; Scholte et al., 2013; Lai et al., 2013), or regions (Clements et al., 2010; Schur et al., 2011a, 2013). However, there are still gaps in mapping areas on a large scale, that is to say continental or global level.

1.5 Objectives of the thesis

The overall goal of this PhD thesis was to develop, implement, and validate Bayesian geostatistical methodology within an epidemiological setting. Focus was placed on the geographical distribution of helminthiases risks (soil-transmitted helminthiasis and schistosomiasis), the ultimate aim being to fill both methodological and epidemiological gaps in order to address the urgent needs of identifying prioritizing areas and estimating the risk and its implications in the frame of control interventions.

1.5.1 Specific objectives

There are seven specific objectives linked to this goal:

(i) review of the major statistical challenges in modelling historical data, together with Bayesian approaches developed to overcome these issues (Chapter 2);

(ii) extraction of historical survey data from published and grey literature pertaining on *A. lumbricoides*, *T. trichiura*, and hookworm for the entire South American subcontinent and on *S. mansoni* for Côte d'Ivoire (Chapters 2, 4–6,);

(iii) development and validation of a multinomial Bayesian geostatistical model allowing modelling of two different parasite species, including co-infection (Chapter 3);

(iv) development of geostatistical model built on a Bayesian geostatistical variable selection allowing selection of non-linear functional form of predictors (i. e. categorical, spline), followed by an assessment of model performance compared to popular selection approaches (Chapter 2);

(v) extension of Bayesian variable selection with prior parameterisation allowing selection among highly correlated predictors, while addressing non-linearity (Chapter 5);

(vi) extension of Bayesian variable selection for large dataset through the estimation of the spatial process *via* the predictive sample approach (Chapter 6); and

(vii) development and validation of Bayesian spatio-temporal geostatistical models allowing the spatial process to vary over time (Chapter 6).

Extracted historical survey data were georeferenced and collated into the open-access GNTD database. With the exception of the multinomial modelling for which data from a recent national Ivorian survey were utilised, the models mentioned above were applied to those historical data:

(i) to identify the most important environmental and socioeconomic predictors and assess their effect on the infections transmission in different settings;

(ii) to produce smooth infection risk maps for the three major soil-transmitted helminth infections at national scale in Bolivia and Brazil, as well as at broader-scale for the entire South-American subcontinent, and for both *S. mansoni* and overall schistosomiasis in Côte d'Ivoire; and

(iii) to derive spatially-explicit estimates of the number of people infected, as well as in need of preventive chemotherapy.

Chapter 2

Statistical methodological issues in mapping historical schistosomiasis survey data

Chammartin F.[1,2], Hürlimann E.[1,2], Raso G.[1,2,3], N'Goran E.K.[3,4], Utzinger J.[1,2] and Vounatsou P.[1,2]

[1] Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, P.O. Box, CH-4002 Basel, Switzerland
[2] University of Basel, P.O. Box, CH-4003 Basel, Switzerland
[3] Centre Suisse de Recherches Scientifiques en Côte d'Ivoire, 01 BP 1303, Abidjan 01, Côte d'Ivoire
[4] Unité de Formation et de Recherche Biosciences, Université Félix Houphouët-Boigny, 22 BP 582, Abidjan 22, Côte d'Ivoire

This paper has been published in *Acta Tropica* 2013, 128(2): 345–352.

Abstract

For schistosomiasis and other neglected tropical diseases for which resources for control are still limited, model-based maps are needed for prioritising spatial targeting of control interventions and surveillance of control programmes. Bayesian geostatistical modelling has been widely and effectively used to generate smooth empirical risk maps. In this paper, we review important issues related to the modelling of schistosomiasis risk, including Bayesian computation of large datasets, heterogeneity of historical survey data, stationary and isotropy assumptions and novel approaches for Bayesian geostatistical variable selection. We provide an example of advanced Bayesian geostatistical variable selection based on historical prevalence data of *Schistosoma mansoni* in Côte d'Ivoire. We include a "parameter expanded normal mixture of inverse-gamma" prior for the regression coefficients, which in turn allows selection of blocks of covariates, particularly categorical variables. The implemented Bayesian geostatistical variable selection provided a rigorous approach for the selection of predictors within a Bayesian geostatistical framework, identified the most important predictors of *S. mansoni* infection risk and led to a more parsimonious model compared to traditional selection approaches that ignore the spatial structure in the data. In conclusion, statistical advances in Bayesian geostatistical modelling offer unique opportunities to account for important inherent characteristics of the *Schistosoma* infection, and hence Bayesian geostatistical models can guide the spatial targeting of control interventions.

2.1 Introduction

The control of schistosomiasis and other neglected tropical diseases is going-to-scale, and hence there is a need for data-driven risk maps to enhance the spatial targeting of control interventions (Brooker et al., 2009a; Hürlimann et al., 2011). The pattern of schistosomiasis transmission is governed by social-ecological systems. Indeed, environmental factors influence the distribution and frequency of intermediate host snails, as well as the intra-snail parasite development, whereas poor socio-economic conditions are associated with higher risks of contracting the disease due to unprotected open water contacts (Aagaard-Hansen et al., 2009; Utzinger et al., 2011). Environmental factors are similar in geographically neighbouring areas, introducing spatial correlation on the risk of schistosomiasis. Geostatistical models take into account correlation and relate disease data with its potential risk factors via a regression equation, which allows prediction of the risk at unobserved locations (Diggle et al., 1998; Raso et al., 2006a). Ignoring the spatial correlation in the data might produce incorrect model estimates (Ver Hoef et al., 2001). Hence, geostatistical modelling has become the method of choice for infectious disease risk profiling (Gemperli et al., 2004; Best et al., 2005; Brooker, 2007). Random effects at the survey sites model spatial dependence via a Gaussian spatial process with a covariance matrix that is defined as a function of distances between pairs of locations. Statistical inference is implemented within a Bayesian framework, which overcomes challenges of estimation of very large number of model parameters via powerful Markov chain Monte Carlo (MCMC) computational algorithms (Gelfand and Smith, 1990).

Bayesian geostatistical model predictions (kriging) over a high resolution grid produce smooth empirical maps depicting the pattern of the infection risk (Diggle et al., 2002). While local maps can guide intervention decisions (Clennon et al., 2004; De Moira et al., 2007), large-scale maps provide useful national assessments for treatment needs and the planning of morbidity control, and ultimately the design of elimination strategies (Clements et al., 2006; Schur et al., 2012, 2013).

In this paper, we provide an overview of the major statistical challenges encountered in mapping historical schistosomiasis data, together with the different Bayesian geostatistical approaches developed to overcome these issues. Next, we emphasise the first and crucial step of each modelling process; a standardised variable selection procedure. We provide an example based on historical survey prevalence data of *Schistosoma mansoni* in Côte d'Ivoire, and illustrate a geostatistical model built via rigorous Bayesian variable selection procedures. Model performance is assessed using a geostatistical model based on predictors,

which are selected using variable selection methods for non-spatial data (Riedel et al., 2010; Pullan et al., 2012). Finally, we discuss future research directions to further improve upon mapping of schistosomiasis risk, which will also be of importance for risk mapping and prediction of other neglected tropical diseases.

2.2 Bayesian approaches for risk profiling

2.2.1 Bayesian computation

Geostatistical model fit requires covariance matrix computations of the Gaussian spatial process. In particular, MCMC implementation involves repeatedly large number of matrix inversions. However, when the number of locations N is large (i. e. > 1000), inversions of the $N \times N$ covariance matrix are computationally expensive, time demanding and practically infeasible. To overcome this well-known "large N" problem, the spatial process has to be estimated on a smaller subset of locations via a spatial process approximation, which arises from the conditional expectation of the original process given its realisation over the sub-sampled locations (Banerjee et al., 2008). These models were applied to predict the risk of schistosomiasis using data from 1800 unique georeferenced survey locations in East Africa and West Africa (Schur et al., 2011a, 2013).

Recently, Rue et al. (2009) developed a faster algorithm with the integrated nested Laplace approximation (INLA). In addition, Gaussian Markov random fields can substitute the spatial process through a stochastic partial differential equation (SPDE) (Lindgren et al., 2011). This new approach considerably increases the speed of fitting large spatial data sets and has been successfully applied for mapping leishmaniasis incidence data in Brazil (Karagiannis-Voules et al., 2013). This new perspective could allow routine spatial analyses of disease data by practitioners.

2.2.2 Heterogeneity of historical survey data

In view of the paucity of readily available spatially explicit epidemiological data, geostatistical modelling often rely on historical survey data extracted from a diversity of sources (e. g. Medline and Ministry of Health reports). Survey design heterogeneities, different sensitivity and specificity of the diagnostic techniques employed and non-uniformity of the population screened are important issues that need to be considered (Brooker et al., 2009b, 2010; Hürlimann et al., 2011).

As preventive chemotherapy guidelines for schistosomiasis and other helminthiases primarily target school-aged children (WHO, 2006), accurate estimation of the risk carried by

this specific age group is of great importance. Indeed, age is a known risk factor for schistosomiasis. For example, the prevalence of *S. mansoni* might reach a peak in 10-year-old children or among adolescents/young adults, depending on the level of endemicity (Warren, 1973; Woolhouse, 1998). Analyses of historical survey data rarely allow to account for age as a predictor, because the data are usually not reported by age. Country-wide age-group alignment factors have been estimated and used to adjust population-based estimates (Schur et al., 2011b, 2013). Immigration-death models can estimate age-specific infection prevalence (Holford and Hardy, 1976). Raso et al. (2007) developed a Bayesian version of the immigration-death model, which was able to characterise the age prevalence curve of *S. mansoni* in a community of the Man area in western Côte d'Ivoire. These models can be applied to align age heterogeneous surveys to a standard age-group, e. g. school-aged children.

WHO recommends the Kato-Katz technique (Montresor et al., 1998) to detect *S. mansoni* and other intestinal helminth infections. Parasite eggs (or larvae) are identified by microscopy on a faecal thick smear. This technique is relatively easy to implement in the field, but is known to have a low sensitivity when performed on a single thick smear, especially for light infections (Bergquist et al., 2009). Using multiple thick smears from a single stool sample or, if resources allow, from multiple stool samples, improves diagnostic accuracy (De Vlas and Gryseels, 1992; Utzinger et al., 2001; Booth et al., 2003; Knopp et al., 2008). Sensitivity and specificity of an indirect haemagglutination assay test were incorporated into a Bayesian geostatistical model to relate the observed *S. japonicum* prevalence with the "true" prevalence in Dangtu, People's Republic of China (Wang et al., 2008). However, schistosomiasis historical survey data often lack detailed information regarding diagnostic protocols, and hence, it is difficult to take into account the diagnostic errors (Schur et al., 2011a, 2013).

2.2.3 Relaxing stationary and isotropy assumptions

Geostatistical models in disease mapping are commonly built under the stationary and isotropy assumptions. Indeed, the spatial correlation is often modelled as an exponential function of distance between locations, irrespective of the direction (isotropy) and the location itself (stationary). Control interventions, improving access to clean water, sanitation and hygiene, ecological transformations and environmental specificities might lead to location-specific spatial correlation, rendering the stationary assumption too strong. Non-stationary models have been developed in various neglected tropical disease applications.

For example, Gemperli (2003) modelled stationary spatial processes that were separated and independent on each other through random Voronoi tessellations. Gosoniu et al. (2009) assumed a mixture of tile-specific separated stationary processes. The aforementioned models have been applied to helminth infection, including schistosomiasis (Raso et al., 2006a; Beck-Wörner et al., 2007).

A key factor for the transmission of schistosomiasis is the presence of intermediate host snails (Stensgaard et al., 2013). Hence, it is not surprising that in many areas, the distribution of schistosomiasis closely matches that of freshwater bodies, or irrigation system (Taylor and Makura, 1985; Steinmann et al., 2006). An attempt addressed to model the spatial correlation through a directional correlation function, considering the direction of river flow and shores, led to improved model-based risk predictions of urogenital schistosomiasis (Schur, 2011).

2.2.4 Bayesian variable selection

Environmental predictors can be derived from various sources, most importantly satellite imagery, weather stations and maps (e. g. water bodies and soil types). The most commonly used predictors for schistosomiasis and other environmentally mediated diseases are precipitation, altitude, land surface temperature (LST), vegetation, soil characteristics and hydrology (Appleton, 1978; Brooker, 2002; Schur et al., 2011a). These predictors are often correlated, and hence regression models fail to correctly identify their effects (Craig et al., 2007). Variable selection procedures are therefore required to build parsimonious and well identifiable models.

Thus far, variable selection within a geostatistical modelling framework has received little attention. It is often carried out separately of the geostatistical model fit and as part of an exploratory analysis. In particular, covariates exceeding a predefined threshold of significance either in a bivariate or a stepwise multiple regression model are selected for the final spatial model (Clements et al., 2006; Riedel et al., 2010). In case of co-linearity, predictors are dropped based on goodness of fit criteria or biological plausibility. However, these approaches ignore the spatial structure of the data, leading to possibly erroneous estimates of the covariate and their significance. In addition, the inclusion or exclusion of a variable depends on the order of parameter entry (or exit), especially when the predictors are correlated.

Over the past 20 years, several Bayesian variable selection approaches have emerged, such as the stochastic search variable selection (George and McCulloch, 1993), the variable

selection sampler of Kuo and Mallick (1998), or the Gibbs variable selection developed by Dellaportas et al. (2002). Recently, these approaches have been applied to geostatistical data (Schur et al., 2011a; Giardina et al., 2012). Moreover, a parameter expanded normal mixture of inverse-gamma (peNMIG) spike-and-slab prior allowing selection of blocks of coefficients has been proposed (Scheipl et al., 2012).

In the next section, we provide details of a handful of geostatistical variable selection procedures. With an example on historical *S. mansoni* survey data from Côte d'Ivoire, we highlight the differences and compare the predictive performance of model outcomes based on a stepwise logistic regression, Gibbs geostatistical variable selection, and peNMIG geostatistical variable selection.

2.3 *S. mansoni* risk profiling for Côte d'Ivoire

2.3.1 Data sources and variables

Survey data on *S. mansoni* infection in Côte d'Ivoire were extracted from the open-access global neglected tropical diseases (GNTD) database (www.gntd.org), which compiles surveys resulting from a systematic review of peer-reviewed and grey literature up to October 2012. Details on the search strategy and in-built quality control features have been described elsewhere (Hürlimann et al., 2011). A total of 244 unique survey locations were included, covering the period from 1970 to 2010.

Nine environmental variables were considered in our analysis, namely (i) day LST; (ii) night LST; (iii) rainfall estimates; (iv) altitude; (v) land cover; (vi) normalised difference vegetation index (NDVI); (vii) soil acidity; (viii) soil moisture; and (ix) distances to freshwater bodies. The latter variable was estimated from the Health Mapper's freshwater map (Schur et al., 2011a). Sources of the variables, together with their spatial and temporal resolution, are summarised in Table 2.1.

2.3.2 Model specification

Variables were either standardised (e. g. day LST, night LST, rainfall, NDVI and soil moisture) or categorised (e. g. soil acidity, landcover and altitude) if they presented a non-linear bivariate association to the logit of the observed *S. mansoni* prevalence. A dummy variable related to the period of the survey was also considered to explore potential temporal trends, distinguishing survey data before 2000 and from 2000 onwards.

Table 2.1: Data sources and properties of the environmental predictors explored to model *S. mansoni* infection risk in Côte d'Ivoire.

Data type	Source	Date	Temporal resolution	Spatial resolution
Day land surface temperature	MODIS/Terra[1]	2000–2011	8 days	1 km
Night land surface temperature	MODIS/Terra[1]	2000–2011	8 days	1 km
Normalised difference vegetation index	MODIS/Terra[1]	2000–2011	16 days	1 km
Land cover	MODIS/Terra[1]	2011	Yearly	1 km
Rainfall	ADDS[2]	2000–2011	10 days	8 km
Altitude	DEM[3]	-	-	1 km
Freshwater bodies	Health Mapper[4]	-	-	-
Soil moisture	WISE3[5]	-	-	10 km
Soil acidity	WISE3[5]	-	-	10 km

[1] Moderate Resolution Imaging Spectroradiometer (MODIS). Available at: https://lpdaac.usgs.gov/ (accessed: October 2012).

[2] Africa Data Dissemination Service (ADDS). Available at: http://earlywarning.usgs.gov/adds/ (accessed: October 2012).

[3] Digital Elevation Model (DEM). Available at: http://eros.usgs.gov/ (accessed: October 2012).

[4] Health Mapper database. Available at: http://gis.emro.who.int/PublicHealthMappingGIS/HealthMapper.aspx (accessed: October 2012).

[5] ISRIC-WISE database (WISE3). Available at: http://www.isric.org/ (accessed: October 2012).

Following the standard geostatistical model formulation, we assumed that the survey data arise from a binomial model, that is $Y_i \sim Bn(N_i, p_i)$, where Y_i are the number of infected individuals out of the total screened N_i at location i, $i = 1, ..., n$ and p_i is the probability of infection, such as, $logit(p_i) = \beta_0 + \sum_{k=1}^{K} I_k \beta_k^T X_{ki} + \phi_i$. X is the design matrix representing the potential predictors k, $k = 1, ..., K$ and ϕ_i are the random effects modelling the spatial process, $\phi = (\phi_1, ..., \phi_n)^T \sim MVN(0, \Sigma)$. The variance-covariance matrix Σ models the geographical correlation via an isotropic exponential correlation function of distance, such as $\Sigma_{lm} = \sigma_{sp}^2 exp(-\rho d_{lm})$, where d_{lm} is the Euclidean distance between locations l and m, σ_{sp}^2 is the geographical variability, and ρ controls the rate of the correlation decay. Geographic dependency (range) is defined as the minimum distance at which spatial correlation between locations is smaller than 5% and is calculated by $3/\rho$.

Typically the regression coefficients β_k are assigned vague prior distribution $\beta_k \sim N(0, \tau^2)$, where τ^2 is a fixed large variance (i. e. 100). There are two main approaches

of variable selection in the Bayesian framework. The Kuo and Mallick (1998) type samplers replace β_k by $\beta_k = I_k\alpha_k$, where I_k is an indicator variable taking value 1 if the kth predictor is included in the model and zero otherwise. I_k and α_k are assumed *a priori* independent with $I_k \sim Be(\pi_k)$ and a vague normal distribution for α_k, that is $\alpha_k \sim N(0, \tau^2)$. π_k specifies the inclusion probability of a predictor and is set to 0.5 in the absence of any prior information. Gibbs variable selection algorithms (Dellaportas et al., 2002) introduce a normal mixture prior distribution for α_k, that is $\alpha_k|I_k \sim I_k N(0, \tau^2) + (1 - I_k)N(0, \upsilon_0\tau^2)$, where υ_0 is a small constant shrinking β_k to zero when the predictor is excluded and τ^2 is fixed as above. A complete review of these approaches is provided by O'Hara and Sillanpää (2009).

Ishwaran and Rao (2005) proposed a spike and slab prior on β_k, which is a scaled NMIG, that is $\beta_k \sim N(0, \tau^2)$, where $\tau^2 \sim I_k IG(a_\tau, b_\tau) + (1 - I_k)\upsilon_0 IG(a_\tau, b_\tau)$ and a_τ, b_τ are fixed parameters of non-informative inverse-gamma distribution. The inclusion probability π_k can either be fixed as above or considered as hyper-parameter with beta prior distribution, $\pi_k \sim Beta(a_\pi, b_\pi)$, allowing for greater amount of flexibility in estimating model size. NMIG selects a predictor at the level of the regression coefficient variance and remove subjectivity in the choice of the coefficient variance, while keeping the selective "shrinkage" property. This approach showed good performance for selection of single coefficients, but is not well appropriate for blocks of coefficients arising in non-linear functional forms (e. g. categorical data and splines). Indeed, MCMC sampling tends to oversample the block around zero and, due to slow convergence, is likely to exclude predictors modelled by blocks of coefficients (Gelman et al., 2008).

To remedy this issue, Scheipl et al. (2012) developed a peNMIG prior, which enables simultaneous inclusion or exclusion of large coefficients batches by improving "shrinkage" properties of the resulting marginal prior. In more detail, let β_{kj} be the regression coefficient for the jth category, $j = 1, ..., d_k$ of the kth predictor. β_{kj} is defined as the product of α_k and ξ_{jk}, where α_k models the overall contribution of the kth predictor and ξ_{jk} estimates the effects of each element β_{kj} of the predictor. A mixture of two Gaussian distributions is assumed for ξ_{jk}, $\xi_{jk} \sim N(m_{jk}, 1)$, $m_{jk} \sim 1/2\delta_1(m_{jk}) + 1/2\delta_{-1}(m_{jk})$ which allows to shrink ξ_{jk} towards $|1|$ (multiplicative identity). A NMIG prior is assigned on α_k as explained above.

2.3.3 Model validation

Models were fitted on a random subset of approximately 80% of the survey locations ($n = 194$) and validated on the remaining locations. Predictive ability of each model was assessed with the mean error (ME) by averaging the differences between the observed *S. mansoni* prevalence and the median of schistosomiasis predictive distributions. Uncertainty was measured by summing the standard deviations (SD) of the predictive distributions. The deviance information criterion (DIC) indicated the goodness of fit of the different models fitted on the full dataset (Spiegelhalter et al., 2002).

2.3.4 Implementation details

Inverse-gamma distribution $\sigma^2_{sp} \sim IG(2.01, 1.01)$ was chosen for the spatial variance σ^2_{sp}, and gamma distribution for the spatial decay spatial decay ρ, $\rho \sim G(0.01, 0.01)$. For the Gibbs variable selection, the hyper-parameter π was set to 0.5 and the variance τ^2 to 100, while priors of the peNMIG variable selection were parameterised as follows: $(a_\tau, b_\tau) = (5, 25)$, $(a_\pi, b_\pi) = (1, 1)$ and $\upsilon_0 = 0.00025$. Vague normal prior distributions were assigned to the β_k's, $\beta_k \sim N(0, 100)$ in the final model fit.

MCMC simulations estimated model parameters of the geostatistical analyses implemented in WinBUGS version 14 (Imperial College and Medical Research Council; London, UK). The variable selections were run with one chain samplers, a burn-in of 50,000 iterations and 50,000 iterations were saved to determine the model with the highest posterior probability. Convergence of models for the validation and the fitting was observed after an average of 50,000 iterations and samples of 500 iterations were then extracted for model parameter estimation. Infection risk was predicted via Bayesian kriging (Diggle et al., 1998) over a grid of 1×1 km spatial resolution (352,869 pixels), covering the entire territory of Côte d'Ivoire, using a self-written Fortran 95 code (Compaq Visual Fortran Professional version 6.6.0). Non-spatial variable selection was carried out according to a stepwise logistic regression with a Wald's $p > 0.1$ exit criterion and a $p \leq 0.05$ entry criterion and was implemented in Stata version 12.1 (Stata Corporation; College Station, USA).

2.3.5 Results

Overall, 244 unique georeferenced *S. mansoni* survey locations were obtained when probing the GNTD database for Côte d'Ivoire for the period 1970–2010. The averaged *S. mansoni* prevalence was 37.7% (SD 28.9%). Figure 2.1 shows the geographical distribution of the historical survey locations and the observed prevalence, stratified by survey period split into two phases; before 2000 and from 2000 onwards.

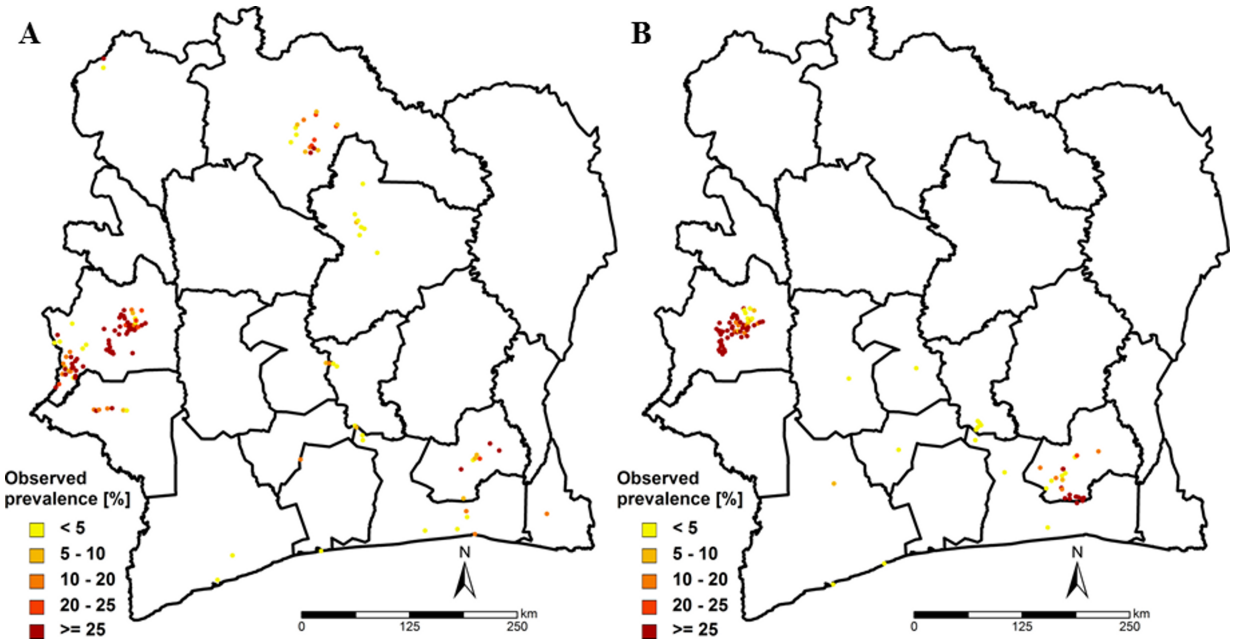


Figure 2.1: Historical survey data of *S. mansoni* prevalence before 2000 (A) and from 2000 onwards (B) in Côte d'Ivoire.

Table 2.2 summarises the results of the variable selection procedures. The stepwise non-spatial logistic regression approach selected nine out of the 10 variables (the distance to water bodies was the only variable excluded). The Gibbs geostatistical approach identified a model including four variables (i. e. period, night LST, rainfall and soil acidity) with posterior probability of 70.0%. The second best model had posterior probability of 16.9% and selected soil moisture instead of rainfall. The three models that fitted the data best, as selected by the peNMIG prior geostatistical approach, were similar to the one selected by the Gibbs geostatistical approach. However, the peNMIG models further included the categorical variable altitude and the models were selected with a lower posterior probability (12.6% for the first model including the period, altitude, night LST, rainfall and soil acidity).

Parameter estimates from the model with the highest posterior probabilities resulting from the stepwise logistic regression (model1), the Gibbs geostatistical variable selection (model 2) and the peNMIG prior geostatistical variable selection (model 3) are given in Table 2.3. Variables selected by the stepwise logistic regression which were not selected by the Bayesian geostatistical approaches had no important effect on the odds of the risk

Table 2.2: Variables selected by a stepwise logistic variable selection and two geostatistical variable selection approaches.

Variable	Stepwise	Gibbs geostatistical			peNMIG geostatistical		
		Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Period	X	X	X	X	X	X	X
Altitude	X	0	0	0	X	0	X
Distance to water bodies	0	0	0	0	0	0	0
Day LST	X	0	0	X	0	0	0
Night LST	X	X	X	X	X	X	X
Land cover	X	0	0	0	0	0	0
Soil acidity	X	X	X	X	X	X	X
Rainfall	X	X	0	X	X	X	0
Soil moisture	X	0	X	0	0	0	X
NDVI	X	0	0	0	0	0	0
Posterior probability (%)		70.0	16.9	4.8	12.6	10.0	6.5

X: selected; 0: not selected; LST, land surface temperature; NDVI, normalised difference vegetation index.
The best three models selected by the geostatistical variable selections are presented, together with their posterior probabilities.

as it is shown by their 95% Bayesian credible interval (BCI). In addition, the model arising from the stepwise logistic regression missed out the effect of rainfall, which was an important predictor in the other two models. The important negative effect of altitude reflected in model 3 was captured neither by the stepwise logistic regression, nor by the Gibbs geostatistical variable selection procedures. The three models revealed comparable goodness of fit properties as reflected by similar DIC measures. However, parameter estimates of model 3 presented less uncertainty, as reflected by a generally lower 95% BCI. In comparison to model 2, model 3 estimated a smaller range (13.3 km *versus* 25.1 km) and a lower variance (3.44 *versus* 3.92), indicating that part of the spatial correlation was captured by the additional variable of model 3, i. e. altitude. In addition, model 1 had a higher variance and a larger spatial range than model 3, suggesting that the additional variables included in model 1 were not able to capture the remaining spatial correlation.

Table 2.4 shows the predictive ability performances of the three models. All models underestimated the observed prevalence, as reflected by the negative ME. Model 3 showed the highest predictive performance, as reflected by the smallest ME (-5.28) and lower uncertainty (overall SD: 7.51). In term of predictive ability, model 1 performed second best while in terms of uncertainty, model 2 was second best.

Table 2.3: Geostatistical logistic regression parameter estimates for *S. mansoni* infection risk summarised by odds ratio (OR) and 95% Bayesian credible intervals (BCI) for the model with predictors selected by a stepwise logistic regression (model 1), a Gibbs geostatistical variable selection (model 2) and a peNMIG prior geostatistical variable selection (model 3).

	Model 1 OR (95% BCI)	**Model 2 OR (95% BCI)**	**Model 3 OR (95% BCI)**
Period			
Before 2000	1.00	1.00	1.00
From 2000 onwards	0.29 (0.24; 0.34)*	0.29 (0.25; 0.34)*	0.29 (0.24; 0.34)*
Soil acidity (pH)			
< 5.20	1.00	1.00	1.00
5.20–5.25	1.36 (0.66; 3.01)	1.65 (0.50; 4.41)	0.97 (0.40; 1.90)
5.25–5.30	1.13 (0.40; 2.31)	1.16 (0.37; 3.34)	0.53 (0.23; 1.34)
≥ 5.30	0.19 (0.05; 0.47)*	0.31 (0.10; 0.85)*	0.14 (0.08; 0.23)*
Night LST	0.41 (0.19; 0.69)*	0.46 (0.33; 0.62)*	0.33 (0.24; 0.47)*
Rainfall	1.15 (0.65; 1.86)	1.59 (1.13; 2.31)*	1.71 (1.25; 2.27)*
Day LST	0.86 (0.52; 1.59)		
NDVI	1.06 (0.67; 1.67)		
Soil moisture	1.44 (0.45; 2.30)		
Land cover			
Urban	1.00		
Evergreen forest	1.38 (0.29; 3.21)		
Deciduous forest/savanna	0.63 (0.10; 2.93)		
Cropland	1.13 (0.26; 2.38)		
Wet areas	0.15 (0.01; 1.89)		
Altitude (m)			
< 270	1.00		1.00
270–340	0.85 (0.48; 2.03)		0.70 (0.57; 0.89)*
≥ 340	0.76 (0.39; 2.05)		0.64 (0.30; 1.09)
Range (km)	14.9 (9.4; 23.3)[1]	25.1 (12.8; 43.5)[1]	13.3 (7.8; 22.7)[1]
Variance	3.59 (2.68; 5.22)[1]	3.92 (2.59; 6.51)[1]	3.44 (2.50; 5.00)[1]
DIC[2]	1,510	1,510	1,508

LST, land surface temperature; NDVI, normalised difference vegetation index.

* Significant based on 95% BCI.

[1] Original scale instead of odds ratio.

[2] The deviance information criterion (DIC) is a measure of goodness of fit. A smaller DIC indicates a better fit to the model data.

S. mansoni infection risk has been predicted on a grid of 1×1 km spatial resolution covering the entire territory of Côte d'Ivoire based on the five predictors identified by variable selection providing the best predictive performance (i. e. the peNMIG prior variable selection). Resulting smooth maps of the estimated risk before 2000 and from 2000 onwards are presented in Figure 2.2. The risk of infection with *S. mansoni* has considerably decreased for the period from 2000 onwards compared to the 30–year period before. High risk areas, where the predicted infection risk exceeds 20%, are found in the region of Man (western part of Côte d'Ivoire), around Lake Kossou (central Côte d'Ivoire), in the district of Azaguié (north of Abidjan) and around the Comoé River (central-eastern part).

Table 2.4: Validation results based on mean error (ME) and sum of the standard deviation (SD) with predictors selected by a stepwise logistic regression (model 1), a Gibbs geostatistical variable selection (model 2) and a peNMIG prior geostatistical variable selection (model 3).

	Model 1	**Model 2**	**Model 3**
ME	-5.41	-5.84	-5.28
Sum of the SD	7.73	7.53	7.51

2.4 Discussion and outlook

Geostatistical models have been widely and successfully used in disease mapping, including schistosomiasis and other neglected tropical diseases (Vounatsou et al., 2009; Machault et al., 2011; Soares Magalhães et al., 2011). Although Bayesian modelling has become the state-of-the-art statistical approach for analysing geostatistical data, there are a number of underlying assumptions that warrant further discussion.

Here, we report advances made to account for four important statistical issues encountered in schistosomiasis risk profiling, namely (i) Bayesian computations involved in fitting large spatial data; (ii) heterogeneities of historical survey data obtained from various sources; (iii) assumptions of stationarity and isotropy; and (iv) variable selection approaches for spatially correlated data. We show that large datasets can be modelled via approximations of the spatial process. Efforts have also been undertaken to account for non-stationarity and anisotropy. However, these models have not been routinely applied because they are highly parameterised and require statistical expertise to develop computationally efficient algorithms and fast processors to obtain estimates within realistic timeframes.

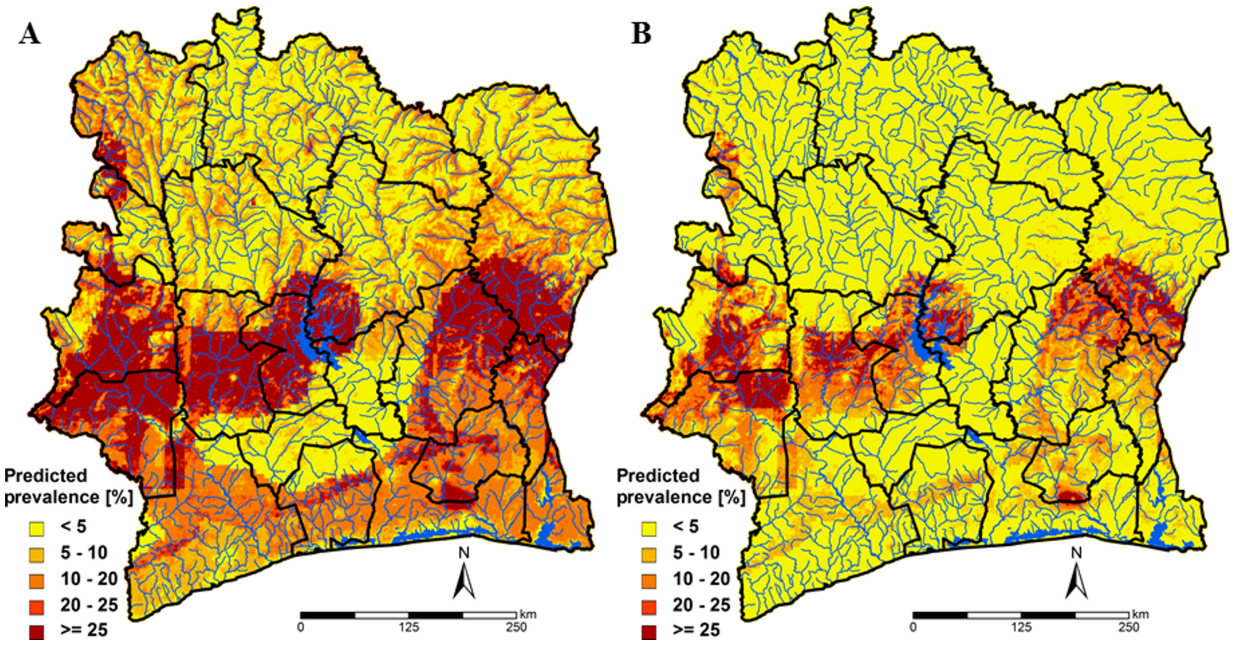


Figure 2.2: Smoothed map of *S. mansoni* infection risk based on predictors selected by a peNMIG Bayesian geostatistical variable selection before 2000 (A) and from 2000 onwards (B) in Côte d'Ivoire.

An important limitation of predictive spatial modelling is the quality of the underlying data. Until contemporary national survey data become available, schistosomiasis risk models will rely on analyses of historical data. One might challenge the comparability of those data, as well as the diagnostic accuracy of the methods used to detect infections, but this goes beyond the methodological scope of the current communication. Geostatistical methodology allows the modelling of diagnostic errors. However, the information reported in the studies that met the inclusion criteria to be integrated into the GNTD database (Hürlimann et al., 2011) often lacked methodological details on the exact diagnostic approach and supply effort. Some authors report prevalence estimates arising from a combination of diagnostic methods. Hence, any effort to adjust for diagnostic sensitivity and specificity is more likely to introduce bias. Age-heterogeneity between surveys is another known problem in geostatistical meta-analyses. Mathematical transmission models have been incorporated in geostatistical modelling to age-align historical malaria survey data (Gemperli et al., 2006; Hay et al., 2009). Similar approaches have yet to be employed in schistosomiasis risk profiling.

The set of predictors included in geostatistical models depends on the variable selection procedures. With our example from Côte d'Ivoire, we demonstrate that a Bayesian geostatistical variable selection with a peNMIG prior selected the predictors which had the most important effects on the *S. mansoni* infection risk and revealed a better predictive model of *S. mansoni* infection risk than previous models. With regard to potential predictors, we included environmental variables that are known to be related to the risk of schistosomiasis (Appleton, 1978; Brooker, 2002; Schur et al., 2011a). Alkaline environment promotes growth and fecundity of freshwater intermediate host snails (Brown, 1994; Southgate, 1997). However, similar to previous findings in East Africa (Schur et al., 2013), pH had a negative effect on *S. mansoni* infection risk. In fact, acid soil is more often found in areas of high rainfall, which might positively affect snail population. In our specific example, the underlying spatial correlation, after accounting for environmental factors, was only moderate (13.3–25.1 km), which explains the small difference in terms of predictive ability of the three resulting models. The maps estimated from the stepwise logistic regression and the Gibbs geostatistical variable selection (results not presented here) showed similar pattern of the *S. mansoni* infection risk with the one arising from the peNMIG geostatistical variable selection. Importantly though, the Bayesian geostatistical variable selection has the advantage of exploring all possible models, while accounting for spatial correlation in the data, which results in more parsimonious models. Accounting for the underlying spatial structure in the selection of predictors should not be ignored, as it influences the estimates of the effect of the covariates and, as a consequence, the selection itself. Inclusion of non-important predictors generates noise and might lead to higher uncertainty.

The Gibbs geostatistical variable selection missed out the important negative effect of the categorical variable altitude, which was captured by the peNMIG prior. Furthermore, the best model identified by the peNMIG prior variable selection presented a much lower posterior probability (12.6% *versus* 70.0%). This shows the good mixing properties of the peNMIG approach and highlights the oversampling around zero of the Gibbs variable selection when exploring inclusion of blocks of predictors. It follows that the peNMIG prior provides more rigorous parameterisation of the regression coefficients, allowing selection of blocks of predictor simultaneously. This parameterisation is of particular interest in disease modelling and we are currently extending this approach to identify the best functional form of a predictor (i. e. linear, categorical or spline) and to select other component of the model, like the spatial random effect which might not be important when spatial correlation is low.

In conclusion, rapid identification of at-risk populations is a key step for cost-effective planning and implementation of disease control intervention. In our view, Bayesian geostatistical models are appropriate tools to obtain predictive maps of disease risk and to estimate the number of infected people in need of treatment and other interventions. Data-driven models, together with recent national surveys, will play important roles in monitoring and impact evaluation of control programmes to reduce the disease burden and ultimately disease elimination.

Acknowledgements

This work received financial support from the European Union grant FP6-STREP-2004-INCO-DEV, project CONTRAST no. 032203. FC is grateful to Foundation of the Centre Suisse de Recherches Scientifiques en Côte d'Ivoire for a PhD fellowship. EH, GR, EKN, JU, and PV are grateful to the Swiss National Science Foundation for financial support (project nos. 320038-132949/1, IZ70Z0-123900, and 325230-118379).

Chapter 3

Bayesian risk mapping and model-based estimation of *Schistosoma haematobium*–*Schistosoma mansoni* co-distribution in Côte d'Ivoire

Chammartin F.[1,2], Houngbedji C.A.[3,4], Hürlimann E.[1,2,3], Yapi R.B.[3,4], Silué K.D.[3,5], Soro G.[6], Kouamé F.N.[6], N'Goran E.K.[3,4], Utzinger J.[1,2], Raso G.[1,2] and Vounatsou P.[1,2]

[1] Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland

[2] University of Basel, Switzerland

[3] Centre Suisse de Recherches Scientifiques en Côte d'Ivoire, Abidjan, Côte d'Ivoire

[4] Unité de Formation et de Recherche des Sciences de la Nature, Université Nangui Abrogoua, Abidjan, Côte d'Ivoire

[5] Unité de Formation et de Recherche Biosciences, Université Félix Houphouët-Boigny, Abidjan, Côte d'Ivoire

[6] Programme National de Santé Scolaire et Universitaire, Abidjan, Côte d'Ivoire

This paper has been formally accepted for publication in *PLoS Neglected Tropical Diseases*.

Abstract

Background: *Schistosoma haematobium* and *Schistosoma mansoni* are blood flukes that cause urogenital and intestinal schistosomiasis, respectively. In Côte d'Ivoire, both species are endemic and control efforts are being scaled up. Accurate knowledge of the geographical distribution, including delineation of high-risk areas, is a central feature for spatial targeting of interventions. Thus far, model-based predictive risk mapping of schistosomiasis has relied on historical data of separate parasite species.

Methodology: We analysed data pertaining to *Schistosoma* infection among school-aged children obtained from a national, cross-sectional survey conducted between November 2011 and February 2012. More than 5,000 children in 92 schools across Côte d'Ivoire participated. Bayesian geostatistical multinomial models were developed to assess infection risk, including *S. haematobium–S. mansoni* co-infection. The predicted risk of schistosomiasis was utilised to estimate the number of children that need preventive chemotherapy with praziquantel according to World Health Organization guidelines.

Principal findings: We estimated that 8.9% of school-aged children in Côte d'Ivoire are affected by schistosomiasis; 5.3% with *S. haematobium* and 3.8% with *S. mansoni*. Approximately 2 million annualised praziquantel treatments would be required for preventive chemotherapy at health districts level. The distinct spatial patterns of *S. haematobium* and *S. mansoni* imply that co-infection is of little importance across the country.

Conclusions/Significance: We provide a comprehensive analysis of the spatial distribution of schistosomiasis risk among school-aged children in Côte d'Ivoire and a strong empirical basis for a rational targeting of control interventions.

3.1 Introduction

The fight against schistosomiasis has been stepped up with global awareness of the burden inflicted upon people who mainly live in rural settings of tropical and sub-tropical countries. Control measures aim to prevent and reduce morbidity due to chronic infection. Whenever resources allow, integrated approaches are advocated that combine preventive chemotherapy targeting school-aged children and other at-risk groups with information, education, and communication (IEC), improvement of sanitation, access to clean water, and focal control of intermediate host snails (WHO, 2002b; Utzinger et al., 2009; Colley et al., 2014). In some countries, long-term concerted efforts successfully controlled morbidity or even achieved interruption of transmission and local elimination (WHO, 2009; Rollinson et al., 2013). However, the World Health Organization (WHO) minimum goal to regularly administer the antischistosomal drug praziquantel to at least 75% of school-aged children at risk of morbidity is far from being reached (i. e. in 2012, coverage in Africa was only 13.6%) (WHO, 2014). Schistosomiasis therefore still remains a major public health concern with a conservative 2010 burden estimated at 3.3 million disability-adjusted life years (DALYs) (Murray et al., 2013).

In Côte d'Ivoire, urogenital and intestinal schistosomiasis are both endemic, caused by chronic infection with *Schistosoma haematobium* and *S. mansoni*, respectively. Efforts to establish a national schistosomiasis control programme date back to the mid-1990s. However, due to the lack of political will and financial resources, and a decade-long socio-political crisis, the programme never really took off (Tchuem Tchuenté and N'Goran, 2009; Bonfoh et al., 2011). In 2010, the "Integrated control of schistosomiasis in sub-Saharan Africa" (ICOSA) project had identified Côte d'Ivoire as a country where preventive chemotherapy is urgently required and should follow WHO guidelines (http://www3.imperial.ac.uk/schisto/wherewework/dfid).

Empirical estimates of the infection risk at the administrative unit where interventions are to be implemented (e. g. health district) are necessary for efficient, cost-effective and sustainable targeting of control measures (Simoonga et al., 2009; Soares Magalhães et al., 2011; Schur et al., 2012). Hierarchical Bayesian geostatistical models provide a robust methodology to establish the statistical relationship between environmental/socioeconomic predictors and the observed risk, while taking into account the spatial dependence inherent to the data. In more detail, it is assumed that the infection risk is driven by a latent spatial Gaussian process, where effects not fully explained by the covariates are captured by a spatial structure in the hierarchy. These models are used in a second step to predict

the risk, including uncertainty, at high spatial resolution using Bayesian kriging methods for spatial process interpolation (Diggle et al., 1998).

Model-based estimates reporting about schistosomiasis risk in Côte d'Ivoire come from single species analyses at district (Raso et al., 2005; Beck-Wörner et al., 2007), national (Chammartin et al., 2013a), or regional level (Schur et al., 2013). Country-wide analyses of schistosomiasis risk are based on historical data that are often heterogeneous (Schur et al., 2013; Chammartin et al., 2013a) and might oversample high endemicity areas as research naturally drives data collection in places where infections are known to be of particular public health concern. Thus, there is a paucity of recent surveys that employed a sampling design that can be utilised for subsequent Bayesian geostatistical analyses of infection risk. Furthermore, the schistosomiasis risk is generally calculated from single species, either using probabilistic laws that assume independence between species (Schur et al., 2011a, 2013), or by applying a correction factor allowing for association between species (De Silva and Hall, 2010; Hodges et al., 2012). However, if the data enable the disease outcome to be categorised into different status of infection (i. e. no, mono-, and co-infection), a geostatistical multinomial model can jointly model the different species (Raso et al., 2006b; Brooker and Clements, 2009).

In the current study, we assessed co-infection risk with both *S. haematobium* and *S. mansoni* and estimated the risk of schistosomiasis in Côte d'Ivoire by analysing recent prevalence data obtained from a national cross-sectional survey conducted in 92 schools across the country (Yapi et al., 2014). We employed a Bayesian geostatistical multinomial model to produce infection risk maps of both *Schistosoma* species, as well as of the overall risk taking into account co-infection. We provide new model-based estimates of the number of infected school-aged children driven by recent data, identify target areas for control measures, and estimate the number of annualised treatments required for deworming the school-aged population.

3.2 Methods

3.2.1 Ethics statement

The study received clearance from the ethics committees of Basel, Switzerland (EKBB, reference no. 30/11) and Côte d'Ivoire (CNER, reference no. 09-2011/MSHP/CNER-P), as well as authorisation from Ivorian Ministry of Education to conduct the study. Prior to the survey, district health and education authorities, school directors, and teachers

were informed about the purpose and procedures of the study. All participants approved verbally their participation to the study and their parents/guardians provided written informed consent. Children infected with *Schistosoma* were treated with a single oral dose of 40 mg/kg praziquantel (WHO, 2002b). In schools where the observed prevalence of schistosomiasis was above 25%, all children were treated with praziquantel regardless of their infection status. Additionally, all children were dewormed with a single dose of 400 mg albendazole (WHO, 2002b).

3.2.2 Study design and survey settings

Details of the study design and survey settings have been described elsewhere (Yapi et al., 2014). In brief, we designed a national cross-sectional survey, combining parasitological examination, clinical observation, and interviewing school children with a questionnaire. The survey was carried out between November 2011 and February 2012 (dry season), just after the country regained political stability after more than 10 years of political unrest (Bonfoh et al., 2011).

Study site selection followed a lattice plus close pairs design (Diggle and Lophaven, 2006). In short, we considered 124 grid cells of 50×50 km overlaid on a map that divides Côte d'Ivoire into two ecological zones: a southern forest area and a northern savannah zone. Ecological zone delimitation resulted from an unsupervised classification via the "iterative self-organizing data analysis technique" (ISODATA) (for more details, see Schur et al. (2011a)). We sampled 54 and 34 grid cells in the southern and northern zone, respectively, proportionally to the population density of the latest available census in 1998. We then randomly selected one locality with a public primary school in each selected grid cell. Six additional school localities were chosen within a 5–20 km radius from the already sampled localities. Teachers of the selected schools were asked to systematically select 60 children attending grades 3–4. If this number was not achieved with classes from grades 3–4, the teachers were asked to select additional children from grade 5. This sample size exceeds the WHO-recommended minimum sample size of 50 for collection of baseline information on helminth prevalence and intensity in the school-aged population within large-scale surveys (WHO, 2006).

3.2.3 Disease data

Study participants were asked to provide a stool and an urine sample. Duplicate Kato-Katz thick smears were prepared shortly after stool collection and examined within 45 min

in situ by two experienced technicians, quantifying *S. mansoni* eggs under a microscope, while microhematuria was assessed using urine using reagent strips (Hemastix®, Bayer, UK) as a proxy for active *S. haematobium* infection. Re-examination of 10% of the slides was performed by senior technicians for quality control.

3.2.4 Environmental, socioeconomic, and population data

Table 3.1 summarises sources and properties of environmental and socioeconomic data investigated to estimate the risk of schistosomiasis in Côte d'Ivoire. In particular, we used satellite-derived estimates such as day and night land surface temperature (LST day and

Table 3.1: Data sources and properties of the variables used to estimate the schistosomiasis risk in Côte d'Ivoire in late 2011/early 2012.

Data type	Source	Temporal resolution	Temporal coverage	Spatial resolution
Day land surface temperature (LST)	MODIS/Terra[1]	8 days	2011	1 km
Night land surface temperature (LST)	MODIS/Terra[1]	8 days	2011	1 km
Normalised difference vegetation index	MODIS/Terra[1]	16 days	2011	1 km
Rainfall	ADDS[2]	10 days	2011	8 km
Altitude	DEM[3]	-	-	1 km
Freshwater bodies	Health Mapper[4]	-	-	-
Soil moisture	WISE3[5]	-	-	10 km
Soil acidity (pH)	WISE3[5]	-	-	10 km
Human influence index (HII)	LTW[6]	-	2005	1 km
Rainfall coefficient of variation (cv)	Derived from rainfall (standard deviation/mean)	10 days	2011	1 km
LST difference	Derived from LST (day LST - night LST)	8 days	2011	1 km
Ecological zone	ISODATA[7]	-	2000–2008	1 km
Improved sanitation index	Bayesian kriging of DHS[8], MICS[9], and WHS[10] with urban/rural[11] as covariate	-	1994–2011	1 km
School-aged population (5–15 years old)	Afripop[12]	-	2010	1 km

[1] Moderate Resolution Imaging Spectroradiometer (MODIS). Available at: https://lpdaac.usgs.gov/ (accessed: October 2012).

[2] Africa Data Dissemination Service (ADDS). Available at: http://earlywarning.usgs.gov/adds/ (accessed: October 2012).

[3] Digital Elevation Model (DEM). Available at: http://eros.usgs.gov/ (accessed: October 2012).

[4] Health Mapper database. Available at:http://gis.emro.who.int/PublicHealthMappingGIS/HealthMapper.aspx (accessed: October 2012).

[5] ISRIC-WISE database (WISE3). Available at: http://www.isric.org/ (accessed: October 2012).

[6] Last of the Wild Project version 2, 2005 (LWP-2): Global Human Influence Index (HII) dataset (geographic) Wildlife Conservation Society International Earth (WCS) and Center for International Earth Science Information Network (CIESIN). Available at: http://sedac.ciesin.columbia.edu/data/set/wildareas-v2-human-influence-index-geographic (accessed: October 2012).

[7] Calculated with the Iterative Self-Organizing Data Analysis Technique (see Schur et al. (2011a))

[8] Demographic and Health Surveys. Available at: http://www.measuredhs.com (accessed: October 2012).

[9] Multiple Indicator Cluster Surveys. Available at: http://www.childinfo.org/mics.html (accessed: October 2012).

[10] World Health Surveys. Available at: http://www.who.int/healthinfo/survey/en/index.html (accessed: October 2012).

[11] Gridded Population of the World version 3. Available at: http://sedac.ciesin.org/gpw/ (accessed: October 2012).

[12] AfriPop version 2.0. Available under request at: http://www.afripop.org (accessed: October 2012).

LST night), normalised difference vegetation index (NDVI), and rainfall estimates. Climatic variation was accounted via the coefficient of variation for rainfall (rainfall cv) and the difference between day and night temperature (LST diff). Soil acidity (pH) and soil moisture expressed supplementary soil characteristics, while additional environmental measures included distance to fresh water bodies and altitude. Ecological zone was accounted as a binary covariate. Socioeconomic proxies were considered with the human influence index (HII) and the percentage of household with improved sanitation (Karagiannis-Voules et al., 2014). The latter was predicted via Bayesian kriging from household survey data collected by the MEASURE Demographic and Health Survey (DHS), the Multiple Indicator Cluster Surveys (MICS), and the World Health Surveys (WHS) programmes. Sanitation facilities were classified as improved following criteria of the Joint Monitoring Programme for Water Supply and Sanitation of WHO and UNICEF (2006). Predictions were adjusted for urban/rural classification and for a binary temporal covariate (trend) with a cut-off at the year 2000. Model-based predictions (of improved sanitation) with and without the temporal trend revealed that the trend term was not important and therefore it was not considered in the predictive model of sanitation. School locations were then overlaid to the resulting kriged surfaces to obtain percentage of household with improved sanitation at survey location. The number of school-aged children (age range 5–15 years) was calculated from the Afripop population density database for the year 2010 and used to estimate the population-adjusted risk and calculate annualised praziquantel treatment needs. In the absence of recent census data (the last census had been done in 1998), we considered the Afripop data as the most accurate estimation of the current population.

3.2.5 Multinomial geostatistical model

The risks of mono-infection with *S. mansoni*, mono-infection with *S. haematobium*, co-infection with the two *Schistosoma* species, and no infection were jointly modelled with a Bayesian multinomial regression model. Spatial correlation was accounted into the model through stationary geostatistical random effects that were assumed to follow a multivariate normal distribution with variance-covariance defined as an exponential function of the distances between any pair of locations. The overall risk of schistosomiasis is then derived by adding up the co-infection risk to the two species-specific mono-infection risks. Similarly, species-specific overall risks are calculated by the sum of the related species mono-infection and the co-infection. Detailed model formulation is given in the Appendix (see subsection 3.5.1).

Bayesian inference of model parameters was performed using Markov chain Monte Carlo (MCMC) simulations in WinBUGS version 14 (Imperial College and Medical Research Council; London, United Kingdom). Models were run with one Gibbs sampler chain for 100,000 iterations and the final 1,000 estimates were used for posterior summaries, validation purposes, and prediction at non-sampled locations. Prediction was carried out at 1×1 km spatial resolution using Bayesian kriging over a grid of more than 350,000 pixels in Fortran 95 (Compaq Visual Fortran Professional version 6.6.0, Compaq Computer Corporation; Houston, United States of America).

3.2.6 Geostatistical variable selection

We performed a geostatistical Gibbs variable selection to identify the most relevant predictors to include in the multinomial geostatistical model (Dellaportas et al., 2002). Our variable selection procedure was run with one Gibbs sampler chain for 100,000 iterations. Posterior inclusion probabilities were calculated on the last 10,000 estimates of each indicator defining the presence or absence of the covariate in the model. Predictors with posterior inclusion probability superior to 50% defined the median probability model (Barbieri and Berger, 2004). Further details on geostatistical variable selection model formulation are provided in the Appendix (see subsection 3.5.2).

3.2.7 Estimated annualised treatment needs

The number of infected school-aged children was calculated for every km^2 by multiplying the predicted prevalence with the number of children aged 5–15 years. As the Ivorian health system is organised in a pyramidal basis with health districts at operational level, the total number of infected children was summed up over health districts and divided by the total population of children to estimate school-aged children adjusted risk. WHO advocates to administer preventive chemotherapy to school-aged children once a year in high endemicity areas (prevalence $> 50\%$), once every 2 years in moderate endemicity areas (10–50%) and twice during primary schooling age in low endemicity areas (WHO, 2006). To calculate treatment needs on a yearly basis, we assumed an average of 6 years of primary schooling and targeted different proportions of the school-aged children population according to the endemicity level (i. e. the entire, half or a third of the population in high, moderate and low endemicity settings, respectively) (Schur et al., 2012).

3.2.8 Model validation

The multinomial geostatistical model was fitted on a random training sample of 72 locations (around 80% of the full dataset). Predictive ability was assessed on the remaining test locations ($L = 20$) with the mean absolute error (MAE) by averaging the absolute differences between predicted $\hat{p}$ and observed prevalences p, such as $MAE = \frac{1}{L}\sum_{i=1}^{20}|\hat{p}_i - p_i|$. Predictive uncertainty was measured by summing the standard deviation (SD) of the predictive distributions.

To validate our multinomial geostatistical approach, we developed additional models under different assumptions. We fitted separate binomial models for each parasite species that assume independence between the infections, as well as a non-stationary multinomial

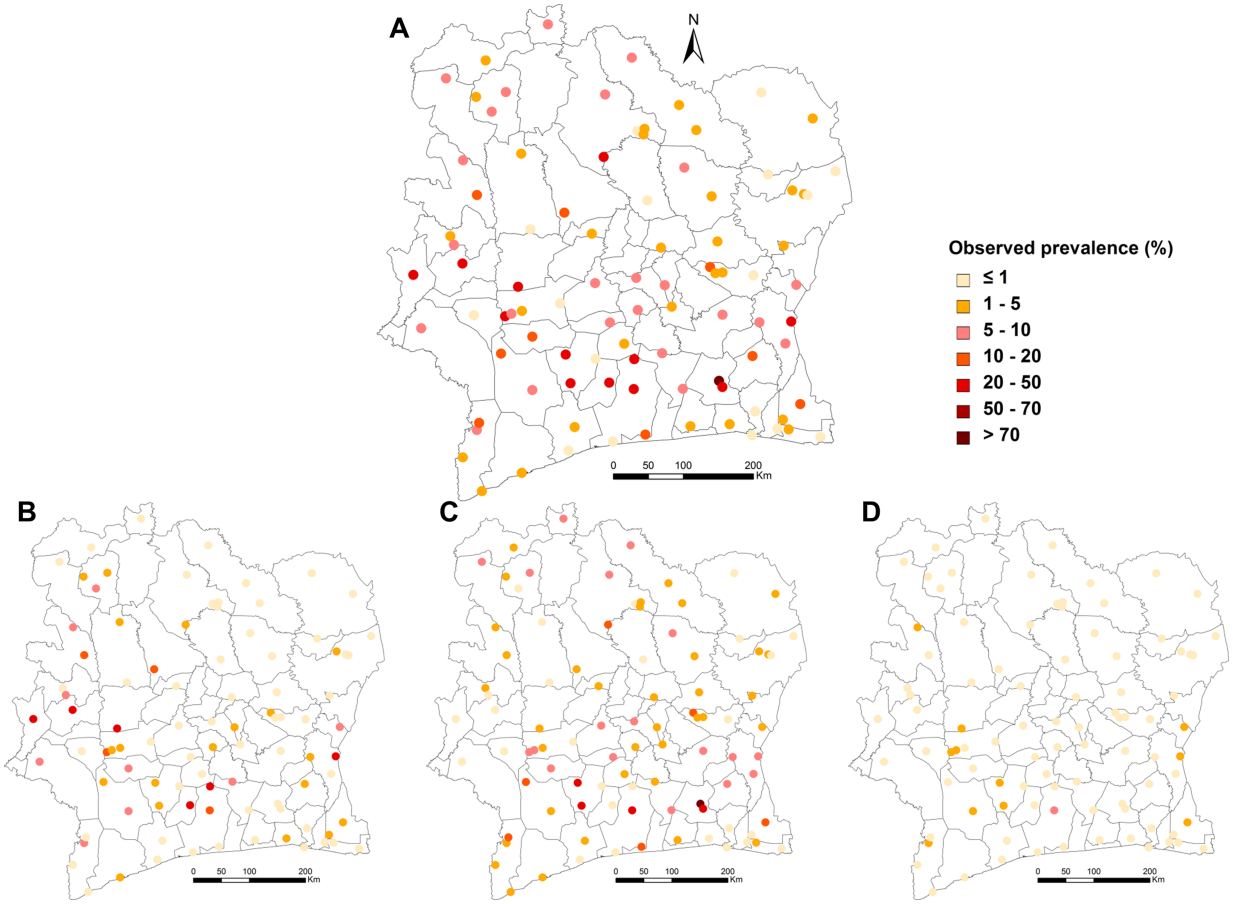


Figure 3.1: Observed schistosomiasis prevalence in Côte d'Ivoire in late 2011/early 2012. A) overall schistosomiasis, irrespective of the species, B) overall *S. mansoni*, C) overall *S. haematobium*, and D) co-infection with both species.

model, which considers that spatial correlation is not only a function of the distances between pairs of locations, but also relies on the locations per se. Thus, we modelled the spatial correlation as a weighted average of ecological zone-specific stationary spatial processes (Beck-Wörner et al., 2007; Gosoniu et al., 2009). Comparison of the predictive ability of those models with our multinomial model was performed in terms of MAE on the overall schistosomiasis risk.

Our prediction were classified according to WHO thresholds for intervention and we compared the observed prevalence of the surveyed schools with the predicted risk at school location, as well as with the school-aged children adjusted risk at health districts level. Number and percentage of schools overestimated and underestimated were calculated to assess the performance of our model-based estimates.

3.3 Results

3.3.1 Disease data

Overall, 5,104 children were examined in 92 schools across Côte d'Ivoire. Out of the 94 schools selected, one school refused to participate and another was excluded since teachers reported deworming interventions during the preceding month. The mean observed prevalence was 5.7% (standard deviation (SD) = 11.2%) for *S. haematobium* and 3.6% (SD = 7.6%) for *S. mansoni* infection. Concomitant infections with both *Schistosoma* species were detected in only 16 children (0.3%, SD = 0.9%), indicating that *S. haematobium*–*S. mansoni* co-infection is rare in Côte d'Ivoire. The spatial distribution of the overall observed prevalence of infection with any *Schistosoma* species is depicted in Figure 3.1, along with the observed distribution of *S. mansoni* and *S. haematobium* single infections, as well as co-infection with both species.

3.3.2 Geostatistical variable selection

Relationships of the 13 potential environmental and socioeconomic predictors with schistosomiasis risk were investigated on the basis of their linear and categorical forms on bivariate non-spatial logistic analyses. Goodness of fit measures showed no benefit to categorise the predictors. Hence, linear predictors were standardised for subsequent analyses. Out of the 13 predictors investigated, LST day was not further considered as the variable was highly correlated to day-night LST difference (correlation coefficient = 0.94). The median probability model, as well as its posterior probability and posterior inclusions

probabilities of the predictors, are presented in Table 3.2. Ecological zone had a high posterior inclusion probability of 93.6%, highlighting the important difference between the two ecological zones regarding the schistosomiasis risk. The median probability model included ecological zone and rainfall coefficient of variation. Furthermore, it was selected among all possible models with the highest posterior probability. The low posterior probabilities of the models explored by our variable selection (below 3.2%), together with the high inclusion probabilities (above 15%) of all potential predictors, suggest good mixing properties of the MCMC simulations and no clear benefit to choose between the explored predictors.

Table 3.2: Geostatistical variable selection results.

Predictors	**Median probability model**	**Predictor posterior inclusion probability**
North ecozone	X	93.6%
Altitude	0	28.9%
Human influence index (HII)	0	15.1%
Soil moisture	0	34.1%
Soil acidity (pH)	0	22.7%
Normalised difference vegetation index	0	15.5%
Night land surface temperature	0	18.4%
Rainfall	0	39.3%
Rainfall coefficient of variation (cv)	X	60.8%
Day-Night difference land surface temperature	0	26.3%
Sanitation index	0	17.4%
Distance to fresh water bodies	0	15.2%
Day land surface temperature	NC	NC
Model posterior probability	3.2%	-

X (selected), 0 (not selected), NC (not considered)
Median probability model is presented together with posterior inclusions probability of the predictors and model posterior probability.

3.3.3 Multinomial geostatistical model

A multinomial logistic model, including ecological zone and rainfall coefficient of variation, was fitted to the data. Estimates of the parameters are presented in Table 3.3, together with predictive ability of the model. Northern savannah ecological zone had a negative effect on the log of the risk of all the multinomial categories *versus* no infection (i. e. *S. mansoni* mono-infection, *S. haematobium* mono-infection, and co-infection with both

Schistosoma species). Higher rainfall variation had a negative effect on *S. haematobium*, and consequently on co-infection, while its effect was not important regarding *S. mansoni* infection risk. Residual spatial correlation was higher for *S. mansoni* mono-infection (153.2 km) than for co-infection risk (107.6 km), and *S. haematobium* mono-infection (66.4 km).

For comparison, we built two additional models; one without predictors and another one with all predictors (parameter estimates and predictive ability results are given in Tables 3.5 and 3.6). The residual spatial correlation was the lowest for each multinomial category in the model with all covariates. This suggests that predictors which have not been selected by the variable selection were able to explain part of the spatial pattern. In addition, our model shows the best predictive ability. While the model including all covariates shows a better MAE regarding *S. mansoni* mono-infection and co-infection with both species, the MAE of the overall schistosomiasis risk is lower. Moreover, our model shows less uncertainty in the predictions as reflected by lower sum of the SD of the posterior predictive distributions at test locations.

Model validation on 20% of observed location also revealed that the multinomial geostatistical model presented in Table 3 predicted better the overall schistosomiasis risk in comparison to a non-stationary multinomial model (MAE: 10.0% *versus* 11.3%), as well as to separated species-specific binomial geostatistical models assuming either independence of the infections (MAE: 10.0% *versus* 11.0%) or dependence accounted through a correction factor (De Silva and Hall, 2010) estimated from the data (MAE: 10.0% *versus* 11.0%; correction factor=0.99).

Smooth map of the overall schistosomiasis risk (*S. mansoni* mono-infection, *S. haematobium* mono-infection and *S. mansoni*–*S. haematobium* co-infection) is depicted in Figure 3.2A. Maps of the risk of infection of *S. mansoni* and *S. haematobium* (mono- and co-infection) are presented in Figure 3.2B and 3.2C, respectively, while the map of co-infection risk alone is shown in Figure 3.2D. We observed that the two species display distinct spatial patterns, which generally do not overlap, and hence, co-infection is low across the country.

Table 3.3: Parameter estimates and predictive ability of Bayesian geostatistical multinomial logistic model.

		S. mansoni **mono-infection**	***S. haematobium*** **mono-infection**	**Schistosomiasis co-infection**
MOR (95% BCI)	North ecozone	0.32 (0.13; 0.99)*	0.39 (0.17; 0.78)*	0.05 (0.01; 0.40)*
	Rainfall coefficient of variation (cv)	0.74 (0.31; 1.47)	0.70 (0.44; 0.99)*	0.37 (0.09; 0.91)*
Median (95% BCI)	Range (km)	153.2 (11.7; 473.9)	66.4 (8.4; 264.2)	107.6 (6.1; 655.1)
	Variance σ^2	5.0 (2.8; 10.4)	1.9 (1.2; 3.7)	1.1 (0.3; 4.2)
Predictive ability (%)	MAE	5.81	6.06	0.57
	Sum of SD	1.58	1.32	0.07

* Significant based on 95% BCI.
Overall schistosomiasis risk: MAE=10.01%, Sum of SD=2.03%.
Multinomial odds ratios (MOR) and median of the spatial parameters estimates are displayed with their 95% Bayesian credible interval (BCI). Predictive ability performance is assessed with a model fitted on a subsample of the data (80%) and is reported by mean absolute error (MAE) and sum of the standard deviation (SD) of the predictive distributions.

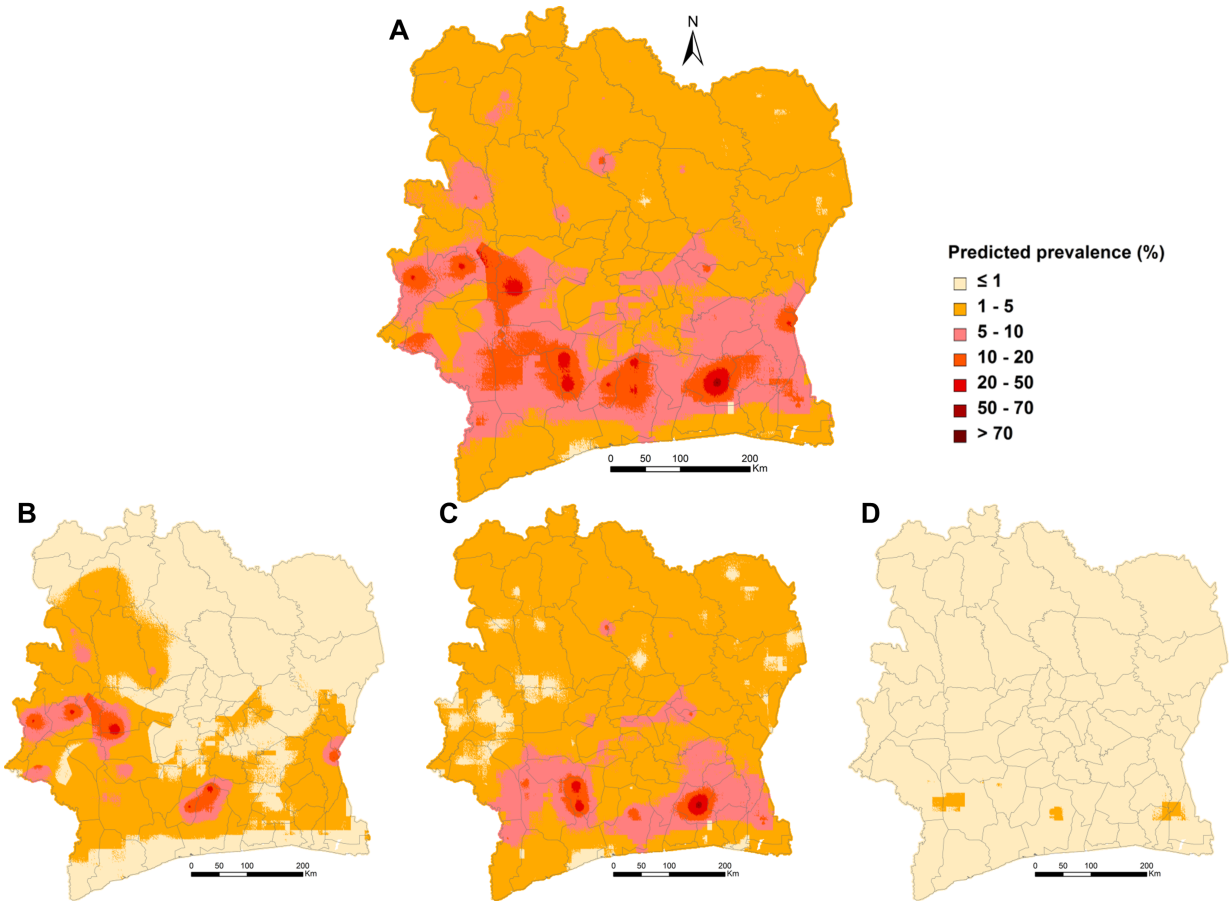


Figure 3.2: Predicted schistosomiasis risk in Côte d'Ivoire in late 2011/early 2012. A) overall schistosomiasis, irrespective of the species, B) overall *S. mansoni*, C) overall *S. haematobium*, and D) co-infection with both species.

3.3.4 Risk and estimated annualised treatment need

In Côte d'Ivoire, we estimated that around 457,062 school-aged children are infected with *Schistosoma*, which correspond to 8.9% of the school-aged population (95% Bayesian credible interval (BCI): 7.5–10.6%; child population aged 5–15 years: 5,135,531). Single species infection risk was estimated at 5.3% (95% BCI: 4.3–6.8%) for *S. haematobium* and 3.8% (95% BCI: 2.9–5.3%) for *S. mansoni*. The children-adjusted risk aggregated at health district level is detailed in Table 3.7. The health district of Agboville presents the highest risk estimated to 29.7%. Health districts were classified as low (predicted children-adjusted risk $< 10\%$) or moderate (predicted children adjusted risk 10–50%) endemic and the resulting map is presented in Figure 3.3. Based on this classification, we calculated that a

total of 1,999,629 annualised praziquantel treatments are required for implementation of preventive chemotherapy against schistosomiasis at health districts level in Côte d'Ivoire. High-risk areas extend in the south-western part of the country, as well as in the northern areas of Abidjan. Misclassification of the surveyed schools by the predicted risk at school (pixel) and health districts levels is provided in Table 4. Our estimates of the schistosomiasis risk misclassify 4.3% of the surveyed schools, while our predictions aggregated at health district level incorrectly classify 22.1% of the visited schools.

Table 3.4: Misclassification of the surveyed schools by the predicted risk at school and health districts level.

School estimated schistosomiasis risk	**$< 10\%$**	**10–50%**	**$\geq 50\%$**
Schools underestimated	4 (4.3%)	-	-
Schools overestimated	-	-	-
Schools misclassified	4 (4.3%)	-	-
Health district estimated schistosomiasis risk	**$< 10\%$**	**10–50%**	**$\geq 50\%$**
Schools underestimated	-	6 (6.5%)	1 (1.1%)
Schools overestimated	-	9 (14.5%)	-
Schools misclassified	-	15 (21.0%)	1 (1.1%)

Number and percentage of schools overestimated and underestimated are given according to endemic thresholds defined by WHO for control interventions.

3.4 Discussion

We present a comprehensive analysis of the spatial distribution of schistosomiasis risk among school-aged children in Côte d'Ivoire. Our predictive map of the overall risk of schistosomiasis confirms that the disease is endemic throughout Côte d'Ivoire and provides a strong empirical basis for rational targeting of preventive chemotherapy and other control measures.

To our knowledge, this is the first estimation of the overall schistosomiasis risk that has been based on a joint analysis of the two *Schistosoma* species that occur in Côte d'Ivoire, taking into account co-infection risk. Our analysis presents further insights compared to previous modelling efforts that have been done in Côte d'Ivoire (Raso et al., 2005; Beck-Wörner et al., 2007; Schur et al., 2011a; Chammartin et al., 2013a). In particular, our predictions are based on recent survey data, where survey locations have been sampled in order to provide an optimal spatial distribution for geostatistical modelling. Although

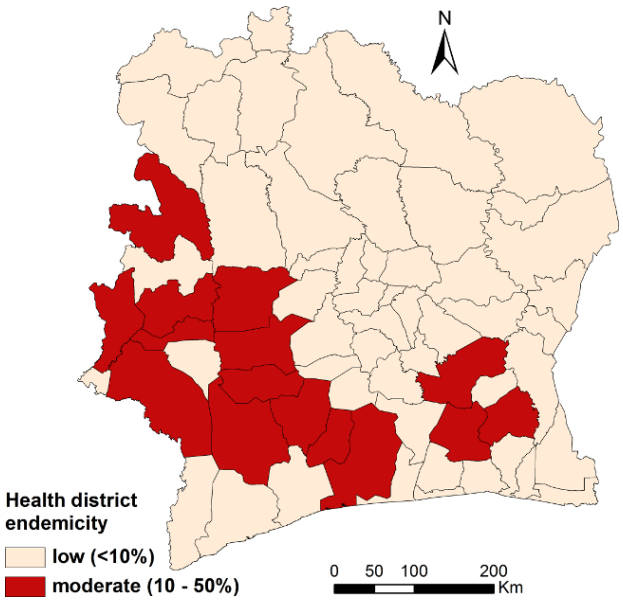


Figure 3.3: Estimated number of school-aged children at risk of schistosomiasis. Maps derived using WHO guidelines and stratified for health districts for control intervention planning.

"lot quality assurance" sampling (Sturrock et al., 2011) has resulted in better predictive performance compared to a geostatistical sampling similar to the one developed in this manuscript, the 92 schools sampled provide a good coverage of the entire surface area of Côte d'Ivoire (322,000 km^2) and a sound basis to quantify the spatial structure of the risk at national level with limited financial resources.

In this study we put forth maps of co-infection rather than co-endemicity risk. The former gives the probability of simultaneous infections at the individual patient level. The latter gives the probability that both infections are present at a given locality. Co-infection implies co-endemicity but not necessarily the other way round. Thus, spatial patterns of co-endemicity and co-infection are not necessarily the same. In fact the definition of co-endemicity in the literature of spatial epidemiology is confusing. In some instances co-endemicity refers to co-infection in others it is calculated as the prevalence of either infection.

We estimated that 8.9% of school-aged children are affected by schistosomiasis in Côte d'Ivoire. This estimate is considerably lower than previously published predictions. For example, Schur et al. (2011a) estimated that 41.8% (95% BCI = 24.4–60.8%) of the population below 20 years of age is infected with schistosomes in Côte d'Ivoire based on an analysis of historical data in West Africa. With regard to *S. mansoni*, our estimate of 3.8%

is also several-fold lower than the previously published prevalence of 11.0% (95% BCI: 8.7–13.8%) that has been calculated based on an analysis of historical data at national level (Chammartin et al., 2013a). Historical data mainly stem from surveys conducted for other purposes than risk mapping and highly endemic areas were likely oversampled. The current analysis therefore highlights the importance of a rigorous sampling design and mapping activities before launching a national control programme. High quality data obtained from surveys well distributed in space are paramount for accurate identification of diseases distribution and efficient use of limited resources for control (Brooker and Clements, 2009; Sturrock et al., 2011). Côte d'Ivoire had not yet begun implementation of preventive chemotherapy at the time of our survey, and hence, it is unlikely to attribute our considerably lower infection prevalence due to control interventions. Artemisinin-based combination therapy (ACT) is freely distributed as a key strategy against malaria in Côte d'Ivoire. The partial activity of ACT against schistosomiasis (Utzinger et al., 2010) might play a role, which is currently difficult to quantify and would deserve further research.

Our study has several limitations and they are offered for discussion. First, schistosomiasis is known to be focally distributed, governed by the presence of humans, specific intermediate host snails, and human-water contact patterns(Kloos, 1995; Lengeler et al., 2002) and the cross-sectional study design of the present study might not capture well this pattern. Aggregating schistosomiasis risk estimates at health district level revealed important misclassification of the schools (22.1%) within the risk thresholds defined by WHO for interventions. Thus, operational and financial advantages that would provide the targeting of interventions at the level of an existing structure, such as the health districts, is challenging due to the focal nature of schistosomiasis. Given the need to better understand the small-scale heterogeneity through additional surveys (Brooker et al., 2009a), the western part of Côte d'Ivoire that is a well-known focus of *S. mansoni* (Doumenge et al., 1987; Raso et al., 2005; Utzinger et al., 2011), has been selected in 2010 for a 5-year randomised intervention study using different treatment schedules against *S. mansoni*, funded through the Schistosomiasis Consortium for Operational Research and Evaluation (SCORE). It will be interesting to analyse the baseline data from the eligibility study that surveyed 263 villages/schools with about 50 children examined for *S. mansoni* in each village/school using duplicate Kato-Katz thick smears. This analysis might fill an important gap of understanding small-scale heterogeneity of *S. mansoni* in this specific region. Second, parasitological analyses were conducted on the targeted population, i. e school-aged children, using WHO-recommended diagnostic techniques for intervention decisions (WHO, 2006).

Our estimates further refine our prior knowledge of the schistosomiasis situation in Côte d'Ivoire. It should be noted, however that the WHO-recommended diagnostic techniques have limitations. For example, it is widely acknowledged that the Kato-Katz technique and the urine-reagent strip tests lack sensitivity, especially in low endemicity settings (Utzinger et al., 2011), while urine-reagent strip tests have additionally low specificity (French et al., 2007; Robinson et al., 2009). As a consequence, it is likely that our data underestimate the infection prevalence due to these diagnostic dilemmas (Bergquist et al., 2009).

An important objective of our study was to assess the co-infection occurrence among Ivorian school-aged children, given that both *S. haematobium* and *S. mansoni* co-exist in the country. Only 16 of the 5,104 children examined were co-infected, suggesting that co-infection is negligible. This result implies that potential synergistic or antagonistic effects of mixed schistosome species infections on morbidity (Koukounari et al., 2010) are of little public health concern in Côte d'Ivoire. The scarcity of co-infection is mainly due to the specific spatial patterns of the two parasitic infections with minimal overlapping of the two species infection risk, as highlighted by the predicted maps, stratified by species. Parameter estimates of models including all investigated covariates show that *S. haematobium* and *S. mansoni* infections proliferate under specific climatic conditions. We attribute these different environmental effects to distinct ecological habitats of *Bulinus* and *Biomphalaria*, the intermediate host snails of *S. haematobium* and *S. mansoni*, respectively (Steinmann et al., 2006).

Towards the end of 2012, the national schistosomiasis control programme, with support of the Schistosomiasis Control Initiative (SCI) has started its activities, emphasising the treatment of school-aged children in high-risk areas, including additional mapping activities launched in December 2013. The current results, along with additional mapping facilitated by an operational research project in the western part of Côte d'Ivoire (sustaining *S. mansoni* control, financially supported by the Schistosomiasis Consortium for Operational Research and Evaluation) and fine-grained national mapping funded through the SCI, have greatly influenced the roll out of the national schistosomiasis control programme. Thus, we believe that with the breadth of recent activities in collecting up-to-date schistosomiasis data and the developed infection risk models for Côte d'Ivoire, great support can be provided to the Ivorian schistosomiasis control programme in their fight against schistosomiasis. Additional concerted efforts will be required to analyse all the data in a timely manner and discuss the findings with the national schistosomiasis control programme manager to guide and spatially target control interventions.

Acknowledgements

We are grateful to the laboratory technicians (Jean K. Brou, Amani Lingué, Mahamadou Traoré, and Sadikou Touré), drivers, school directors and teachers, and all school children for their participation. We thank Dimitrios-Alexios Karagiannis-Voules for Bayesian kriging of socioeconomic proxies. FC is grateful to Foundation of the Centre Suisse de Recherches Scientifiques en Côte d'Ivoire for a PhD fellowship. The work received financial contribution from the Swiss National Science Foundation (project no. PDFMP3-137156 and 32003B-132949).

3.5 Appendix

3.5.1 Multinomial geostatistical model

We define Y_{jk}, n_j, and p_{jk} as the number of infected, the number of screened, and the probability of mono-infection with *S. mansoni* ($k = 1$), mono-infection with *S. haematobium* ($k = 2$), co-infection with the two species ($k = 3$), and no infection ($k = 4$) at observed location j, $j = 1, ..., 92$. The sum of mono-infections and co-infection ($\sum_{k=1}^{3} p_{jk}$) represents the overall schistosomaisis risk. Assuming that Y_{jk} is multinomially distributed, $Y_{jk} \sim MN(n_j, p_{jk})$, we modelled the log odds of the infection in each multinomial category l, $l = 1, ..., 3$, *versus* the baseline category (no infection) as a linear function of covariates X_j and latent spatially structured Gaussian process φ_{jl} such as: $log(p_{jl}/p_{j4}) = X_j^T \underline{\beta}_l + \varphi_{jl}$, where p_{jl}/p_{j4} is the risk ratio between the infection status and no infection and $\underline{\beta}_l$ is the vector of regression coefficients for each multinomial category l. The spatially structured random effects $\underline{\varphi}_l$'s are modelled as follows: $\underline{\varphi}_l \sim MVN(\underline{0}, \Sigma_l)$, with variance-covariance matrix Σ_l. Spatial dependency is introduced via an exponential correlation function of Euclidian distances between observed locations, i. e. $\Sigma_{rsl} = \sigma_l^2 exp(-\rho_l dist_{rs})$, where σ_l^2 is the geographical variability, ρ_l controls the rate of the spatial decay and (r,s) is a pair of locations. The range indicates the minimum distance at which spatial correlation between locations is less than 5% and is equal to $-log(0.05)/\rho$. Inference was drawn within a Bayesian framework with non-informative prior specification for model parameters. In particular, vague normal distribution was chosen for the regression coefficients, $\underline{\beta} \sim N(\underline{0}, 100)$, inverse-gamma distribution $\sigma^2 \sim IG(2.01, 1.01)$ were adopted for the spatial variance σ^2, and gamma distribution for the spatial decay ρ, $\rho \sim G(0.01, 0.01)$.

3.5.2 Geostatistical variable selection

For each potential covariates m, $m = 1, ..., 12$, we introduced a binary indicator I_m suggesting the presence or absence of the mth variables in the model and a coefficient α_{ml} indicating the effect size of the mth predictors with respect to the lth multinomial category. Each regression coefficient β_{ml} was seen as the product of the indicator I_m and its effect α_{ml}, such that $\beta_{ml} = I_m \alpha_{ml}$. We assumed *a priori* dependence of I_m and α_{ml}, i. e. $P(\alpha_{ml}, I_m) = P(\alpha_{ml}|I_m)P(I_m)$ and assigned to I_m a Bernoulli distribution, such as, $I_m \sim Be(\pi_m)$ with hyper-parameter π_m having a non-informative Beta distribution; $\pi_m \sim Beta(1, 1)$. A mixture normal distribution was assumed for α_{ml}, such as, $P(\alpha_{ml}|I_m) \sim I_m N(0, \tau_m^2) + (1 - I_m) N(0, \tau_m^2/\lambda)$, where τ_m^2 is a variance parameter with a non-informative

inverse-gamma prior distribution, $\tau_m^2 \sim IG(5, 25)$ and λ is a constant set to 4,000 to shrink the α_m's to 0 when the variable X_m is excluded from the model.

Table 3.5: Parameter estimates of Bayesian geostatistical multinomial logistic model including all considered predictors.

		S. mansoni **mono-infection**	***S. haematobium*** **mono-infection**	**Schistosomiasis co-infection**
MOR (95% BCI)	North ecozone	0.05 (0.01; 0.35)*	0.40 (0.22; 1.9)	0.01 (0.00; 0.14)*
	Altitude	2.78 (1.21; 9.23)*	0.66 (0.41; 1.55)	3.33 (0.37; 56.09)*
	Human influence index (HII)	1.42 (0.64; 2.13)	1.01 (0.65; 1.51)	0.61 (0.08; 1.97)
	Soil moisture	2.96 (1.35; 10.71)*	0.83 (0.48; 1.54)	2.82 (0.38; 215.94)
	Soil acidity (pH)	1.66 (0.42; 3.41)	1.62 (0.89; 3.09)	0.82 (0.02; 9.75)
	Normalised difference vegetation index	0.66 (0.35; 1.91)	1.25 (0.62; 2.36)	1.05 (0.21; 4.45)
	Night land surface temperature	0.46 (0.21; 0.79)*	1.01 (0.55; 1.66)	0.72 (0.20; 2.63)
	Rainfall	0.50 (0.22; 1.09)	1.32 (0.79; 2.12)	1.4 (0.11; 7.35)
	Rainfall coefficient of variation (cv)	0.82 (0.51; 1.71)	0.59 (0.37; 0.90)*	0.20 (0.03; 1.22)
	Day-Night difference land surface temperature	0.74 (0.27; 2.29)	0.95 (0.61; 1.70)	2.01 (0.21; 11.53)
	Sanitation index	1.61 (0.93; 2.52)	1.48 (0.83; 2.50)	2.72 (0.91; 5.67)
	Distance to fresh water bodies	0.62 (0.26; 1.12)	0.97 (0.64; 1.32)	1.07 (0.40; 3.14)
Median (95% BCI)	Range (km)	28.9 (5.8; 126.2)	46.3 (7.0; 224.1)	79.9 (6.6; 389.0)
	Variance σ^2	5.1 (2.8; 9.2)	2.2 (1.3; 4.1)	2.9 (0.4; 42.9)
Predictive ability (%)	MAE	5.71	6.28	0.54
	Sum of SD	2.75	1.78	0.57

* Significant based on 95% BCI.

Overall schistosomiasis risk: MAE=10.67%, Sum of SD=3.35%.

Multinomial odds ratios (MOR) and median of the spatial parameters estimates are displayed with their 95% Bayesian credible interval (BCI). Predictive ability performance is assessed with a model fitted on a subsample of the data (80%) and is reported by mean absolute error (MAE) and sum of the standard deviation (SD) of the predictive distributions.

Table 3.6: Parameter estimates of Bayesian geostatistical multinomial logistic model without covariates.

		S. mansoni **mono-infection**	***S. haematobium*** **mono-infection**	**Schistosomiasis co-infection**
Median (95% BCI)	Range (km)	139.8 (16.5; 372.7)	82.1 (18.9; 207.9)	194.7 (10.0; 624.7)
	Variance σ^2	4.1 (2.2; 7.7)	2.0 (1.3; 3.5)	1.8 (0.4; 6.4)
Predictive ability (%)	MAE	5.85	6.96	1.43
	Sum of SD	7.21	8.16	7.28

Overall schistosomiasis risk: MAE=41.92%, Sum of SD=14.73%.
Median of the spatial parameters estimates are displayed with their 95% Bayesian credible interval (BCI).
Predictive ability performance is assessed with a model fitted on a subsample of the data (80%) and is reported by mean absolute error (MAE) and sum of the standard deviation (SD) of the predictive distributions.

Table 3.7: Overall schistosomiasis risk adjusted for school-aged population (5–15 years old), by health disctircts.

Health district	Population	Adjusted risk (%)	Health district	Population	Adjusted risk (%)	Health district	Population	Adjusted risk (%)
Agboville	104828	29.7 (13.9; 40.4)	Duekoue	48001	9.2 (5.4; 13.9)	Koun-Fao	47540	6.0 (2.9; 9.9)
Man	120509	19.2 (5.9; 27.7)	Aboisso	112209	9.1 (5.9; 13.1)	Bouaflé	90364	5.7 (3.4; 9.3)
Divo	200712	17.1 (10.4; 22.3)	Toulepleu	12941	8.9 (4.5; 18.8)	San-Pedro	44782	5.6 (3.0; 10.1)
Danané	102695	16.2 (5.9; 25.0)	Toumodi	42991	8.7 (5.7; 13.2)	Katiola	47918	5.4 (3.4; 8.8)
Vavoua	102100	16.0 (7.7; 25.3)	Dabou	72090	8.4 (5.2; 12.5)	Korhogo	163577	5.3 (3.0; 9.1)
Gagnoa	149079	15.8 (9.8; 22.3)	Didievi	22217	8.3 (5.0; 12.7)	Jacqueville	18968	5.2 (2.7; 9.6)
Adzopé	75130	15.6 (9.7; 22.0)	Sinfra	63715	8.3 (5.5; 13.4)	Bouaké-Sud	4807	5.0 (2.6; 9.5)
Issia	89248	14.9 (10.5; 20.1)	M'Bahiakro	41183	8.2 (5.0; 12.4)	Bouaké-Est	198225	4.8 (2.4; 8.3)
Lakota	54978	14.7 (8.7; 20.6)	Yamassoukro	105453	8.0 (4.2; 13.2)	Grand-Bassam	26643	4.8 (2.0; 9.8)
Soubré	58814	12.7 (9.0; 18.0)	Dimbokro	38873	7.9 (4.2; 12.6)	Sakassou	26673	4.8 (2.8; 8.3)
Abengourou	115722	12.6 (9.6; 17.0)	Tiébissou	25809	7.9 (4.7; 12.3)	Minignan	19745	4.6 (2.3; 8.3)
Sikensi	62618	12.2 (8.6; 17.8)	Oumé	69490	7.6 (4.8; 11.7)	Ferkessédougou	87510	4.4 (2.4; 7.7)
Akoupé	46219	12.1 (7.6; 18.3)	Madinani	15332	7.1 (4.5; 10.9)	Kounahiri	8932	4.4 (2.2; 8.2)
Bangolo	44581	12.1 (5.0; 19.8)	Mankono	50697	7.1 (4.6; 11.2)	Bouna	55176	4.2 (2.2; 7.1)
Guiglo	59828	11.0 (7.3; 16.4)	Tabou	13161	7.0 (4.7; 10.6)	Dabakala	34055	4.2 (2.5; 7.7)
Bongouanou	114827	10.8 (7.3; 15.7)	Daoukro	53769	6.8 (3.8; 11.7)	Abidjan zone	1048866	4.1 (1.2; 8.8)
Daloa	186980	10.5 (6.6; 14.9)	Zuénoula	54582	6.6 (4.4; 10.4)	Bondoukou	145400	4.1 (2.1; 7.3)
Touba	37850	10.0 (5.9; 15.9)	Sassandra	10398	6.3 (3.8; 10.6)	Tengréla	21658	4.1 (1.7; 8.5)
Agnibilekro	42481	9.6 (5.6; 14.3)	Alepe	56547	6.2 (3.4; 10.1)	Nassian	10708	4.0 (1.6; 9.1)
Grand-Lahou	30082	9.6 (6.1; 15.9)	Odienné	33521	6.1 (3.7; 10.1)	Bouaké-Ouest	17957	3.9 (2.0; 6.9)
Biankouma	39013	9.2 (5.0; 14.7)	Séguéla	48981	6.1 (3.2; 11.2)	Beoumi	40392	3.7 (2.0; 6.6)
Bocanda	39701	9.2 (6.2; 13.2)	Boundiali	56277	6.0 (3.4; 10.2)	Adiaké	49400	3.2 (1.3; 6.6)

Adjusted risks are given with their 95% Bayesian credible intervals (BCI) and are listed in decreasing order of importance.

Chapter 4

Soil-transmitted helminth infection in South America: a systematic review and geostatistical meta-analysis

Chammartin F.[1,2], Scholte R.G.C.[1,2,3], Guimarães L.H.[4], Tanner M.[3,4], Utzinger J.[1,2] and Vounatsou P.[1,2]

[1] Swiss Tropical and Public Health Institute, Basel, Switzerland
[2] University of Basel, Basel, Switzerland
[3] Centro de Pesquisas René Rachou, Fiocruz/MG, Barro Preto, Belo Horizonte, Brazil
[4] Immunology service, Hospital Edgard Santos, Federal University of Bahia, Salvador, Bahia, Brazil

This paper has been published in *The Lancet Infectious Diseases* 2013, 13(6): 507–518.

Summary

Background: The four common soil-transmitted helminth species — *Ascaris lumbricoides*, *Trichuris trichiura*, and the two hookworm species *Ancylostoma duodenale* and *Necator americanus* — are endemic in South America, but their distribution, infection prevalence, and regional burden are poorly understood. We aimed to estimate the risk and number of people infected with *A. lumbricoides*, *T. trichiura*, and hookworm across South America.

Methods: We did a systematic review of reports on the prevalence of soil-transmitted helminth infection in South America published up to May 14, 2012. We extracted and georeferenced relevant survey data and did a meta-analysis of the data to assess the geographical distribution of the infection risk with Bayesian geostatistical models. We used advanced Bayesian variable selection to identify environmental determinants that govern the distribution of soil-transmitted helminth infections.

Findings: We screened 4085 scientific papers and identified 174 articles containing relevant survey prevalence data. We georeferenced 6948 survey locations and entered the data into the open-access Global Neglected Tropical Diseases database. Survey data were sparse for the south of the continent and for the western coast, and we identified no relevant information for Uruguay and little data for smaller countries such as Suriname, Guyana, French Guiana, and Ecuador. Population-adjusted prevalence of infection with *A. lumbricoides* was 15.6%, with *T. trichiura* was 12.5%, and with hookworm was 11.9% from 2005 onwards. Risks of contracting soil-transmitted helminth infection have substantially reduced since 2005 (odds ratio 0.47 [95% Bayesian credible interval 0.46–0.47] for *A. lumbricoides*, 0.54 [0.54–0.55] for *T. trichiura*, and 0.58 [0.58–0.59] for hookworm infection).

Interpretation: Our findings offer important baseline support for spatial targeting of soil-transmitted helminthiasis control, and suggest that more information about the prevalence of soil-transmitted helminth infection is needed, especially in countries in which we estimate prevalence of infection to be high but for which current data is scarce.

4.1 Introduction

More than 5 billion people are at risk of soil-transmitted helminthiasis — parasitic worm infections caused by the roundworm *Ascaris lumbricoides*, the whipworm *Trichuris trichiura*, and the two hookworm species *Ancylostoma duodenale* and *Necator americanus* (Pullan and Brooker, 2012). More than 1 billion people worldwide are infected with one or several species of soil-transmitted helminths, with the highest prevalences reported in Africa, Asia, and Latin America (Bethony et al., 2006; Knopp et al., 2012). Chronicity, recurrence, and infections with several species are common in developing countries (Ezeamama et al., 2005; Steinmann et al., 2008). Infections can cause diarrhoea, abdominal pain, weakness, and general ill-health. More serious long-term effects include malnutrition, physical and intellectual growth retardations in children, and reduced work productivity in adults (Bethony et al., 2006; Brooker et al., 2006; Hotez et al., 2008a; Knopp et al., 2012). Soil-transmitted helminth infections are closely connected with poverty (Hotez, 2008) and are endemic in the whole of Latin America and the Caribbean (PAHO and WHO, 2009). Preliminary estimates report that in this part of the world roughly 100 million people are infected with *T. trichiura*, 84 million with *A. lumbricoides*, and 50 million with hookworm (De Silva et al., 2003).

Preventive drug therapy and, whenever resources allow, improved access to clean water together with adequate sanitation and hygiene make up the global strategy to control soil-transmitted helminthiasis (WHO, 2010). Spatially explicit risk maps at the appropriate resolution to plan interventions are needed for the most rational and cost-effective targeting of control. Such maps need survey data and empirical models to predict risk areas, including uncertainty, at locations where actual data might be missing. Bayesian geostatistical models (Diggle et al., 1998) are the most rigorous statistical approaches and such models are increasingly being used to estimate infection risk, number of people infected, and disease burden at high spatial resolutions (Soares Magalhães et al., 2011; Schur et al., 2013). However, the shortage of readily available data is a serious impediment that has delayed mapping of soil-transmitted helminthiasis in Latin America (IADB, PAHO and SABIN, 2011).

We aimed to produce high-resolution estimates of the risk and number of people infected with *A. lumbricoides*, *T. trichiura*, and hookworm across South America by doing a geostatistical meta-analysis of the compiled data using Bayesian approaches and advanced variable selection methods to identify the most important disease predictors.

4.2 Methods

4.2.1 Search strategy and selection criteria

We did a systematic review in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Moher et al., 2009) to identify all relevant publications pertaining to prevalence data on *A. lumbricoides*, *T. trichiura*, and hookworm infections in 13 South American countries (Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, French Guiana, Guyana, Paraguay, Peru, Suriname, Uruguay, and Venezuela). We searched PubMed and Web of Knowledge (covering all dates from the creation of each database up to May 14, 2012) with pertinent keywords related to parasite species and countries (Appendix 4.5.1). We assessed titles and, if available, abstracts and examined relevant reports in full for soil-transmitted helminth prevalence survey data (for main exclusion criteria, see Appendix 4.5.1). No restrictions were placed on date, study design, and language of publication. If surveys were done at the same location at different times, we made sure that they were separated by at least 12 months to avoid bias due to anthelmintic treatment outcomes. We obtained additional data from national Ministry of Health reports and from personal communication with the Pan American Health Organization (PAHO).

4.2.2 Data extraction

We extracted prevalence data from reference sources in accordance with a standard protocol (Appendix 4.5.1). We retrospectively georeferenced the data with online sources of information and entered them into the free open-access Global Neglected Tropical Diseases (GNTD) database so the data can be used by researchers and disease control managers (Hürlimann et al., 2011). We assessed 30% of the publications identified through the systematic literature search for relevance (every third reference listed by author in alphabetical order) and assessed all relevant documents for quality of extraction. We double-checked all coordinates of survey locations in Google Maps to ensure that they referred to a human settlement. We carefully examined publications from the same investigators reporting survey data at the same location and period to avoid duplicate information.

4.2.3 Environmental and population data

We obtained 19 bioclimatic variables from WorldClim Global Climate Data (Hijmans et al., 2005). These variables were derived from spatial interpolations of monthly average climate data from 1950 to 2000. Remotely sensed climatic proxies are not available before

1980. Hence, on the assumption that climate changes in South America were not severe enough to affect the parasites, we used WorldClim data, which cover similar periods to the disease data. We also extracted altitude data from the WorldClim database and variables about the soil acidity and soil moisture from ISRIC World Soil Information (Batjes, 2006). The spatial distributions of the variables used in the final models are presented in the Appendix 4.5.2. We obtained population densities for 2010 from Gridded Population of the World, version 3 (CIESIN, FAO and CIAT, 2005).

4.2.4 Geostatistical meta-analysis

We did a meta-analysis of the compiled survey data to develop predictive geostatistical logistic regression models estimating the risk of *A. lumbricoides*, *T. trichiura*, and hookworm infections at high spatial resolution. Environmental predictors were standardised or treated as categorical if they had a non-linear association with the infection outcome. We introduced a temporal trend as a binary variable, indicating whether a survey was done before or after 2005. Predictors were grouped together if they were highly correlated (i. e. Pearson's correlation coefficient > 0.9). We did Bayesian Gibbs variable selection (Dellaportas et al., 2002) to assess which were the most important of all 22 predictors. The procedure enabled several models to be fitted simultaneously. The models included all possible combinations of predictors, but only allowed a maximum of one predictor from each group of highly correlated predictors. We used the predictors included in the model with the highest posterior probability to fit the final geostatistical predictive model. Spatial correlation was regarded as an exponential function of distance between any pair of locations. The large number of survey locations gave rise to large correlation matrices, complicating model fit. To overcome this computational challenge we did predictive process approximations (Banerjee et al., 2008), estimating the spatial correlation parameters from a subset of 200 locations (knots). We used minimax space filling sampling to select the knots (Johnson et al., 1990; Diggle and Lophaven, 2006). Bayesian kriging (Cressie, 1990) was done to predict the infection risk over a grid of 614 044 pixels at 5×5 km spatial resolution covering South America. The models of *A. lumbricoides*, *T. trichiura*, and hookworm infections were fitted on a random subset of about 80% of the survey locations and validated on the remaining 20% by comparing the observed prevalence data with the model-based estimates. Based on experience, this sample is large enough for our purposes without compromising the amount of available data for model fit. Further details of the variable selection, model fit, and validation are given in the Appendixes 4.5.3 and 4.5.4.

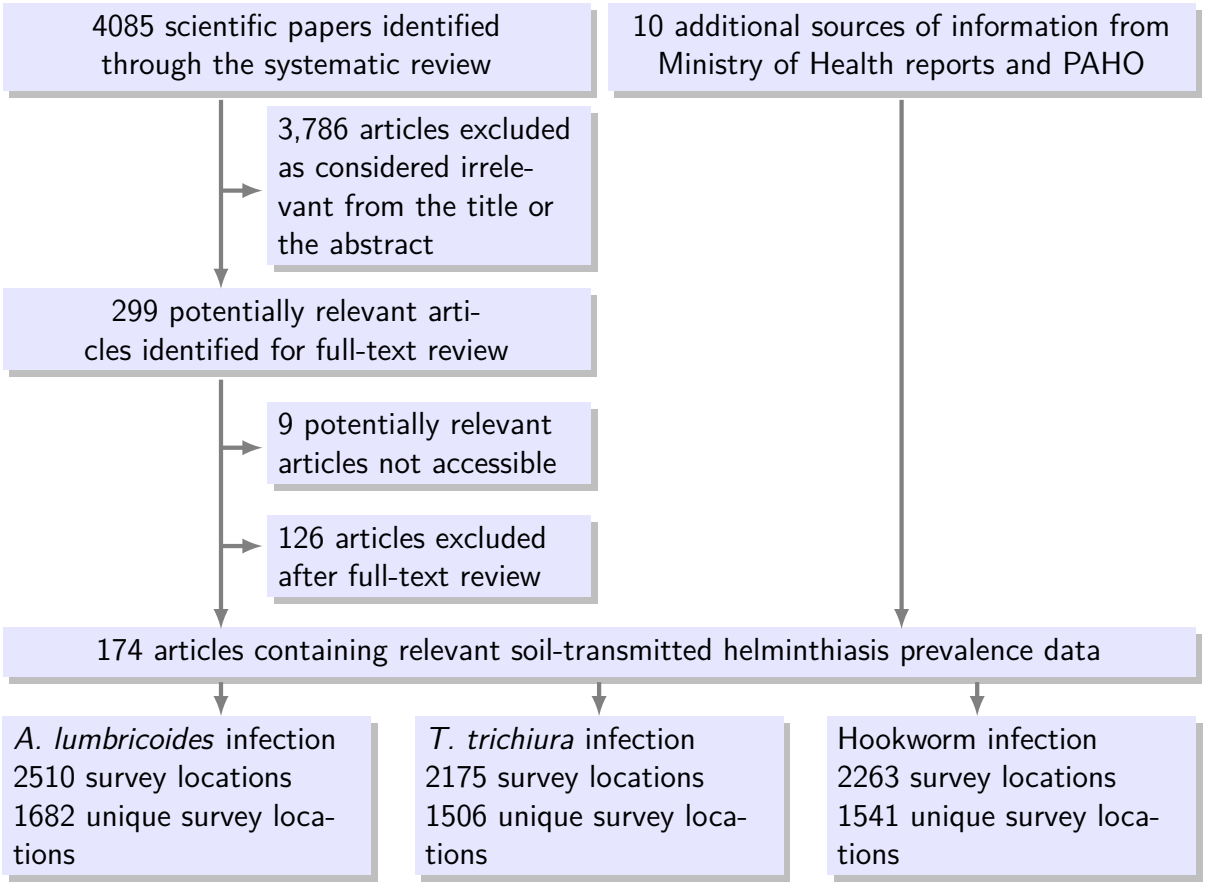


Figure 4.1: Study selection. Survey data were from 1952 to 2011; more than 80% of the data were obtained after 2000.

4.2.5 Role of the funding source

The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The first and corresponding authors had full access to all the data and had final responsibility for the decision to submit for publication.

4.3 Results

Our systematic review identified 4085 peer-reviewed reports. Together with ten additional sources of information from ministries of health and PAHO, 174 references had relevant soil-transmitted helminth infection prevalence data (Figure 4.1). 86% of the survey locations that we identified were associated with survey data reported in grey literature only (ministries of health and PAHO).

An overview of the number of identified surveys with relevant soil-transmitted helminth infection prevalence data, stratified by country, soil-transmitted helminth species, survey period, and geographical unit (i. e. municipality or specific location [point]), is presented in Table 4.1. Most data were from Brazil. Survey data were sparse for the south of the continent, especially for Argentina and Paraguay, and for the western coast (i. e. Chile and Peru; Figures 4.2–4.4). We identified no relevant information for Uruguay. Smaller countries such as Suriname, Guyana, French Guiana, and Ecuador contributed few data, some dating back to 1963. Mean infection prevalences calculated from the observed survey data were 17.0% for *A. lumbricoides*, 9.3% for *T. trichiura*, and 9.4% for hookworm. Most (78%) of the surveys were done after 2000, and 98% were community based, with 95% using the WHO-recommended Kato-Katz technique for diagnosis (Appendix 4.5.6) (Montresor et al., 1998). A summary of the number of surveys with incomplete information, affecting the assessment of the observed prevalence, is provided in the Appendix 4.5.7. Missing description of diagnostic methods (11%) was the most common reason for incomplete information.

Results of the variable selection to identify the most important environmental predictors of disease are given in Table 4.2. The best models had high posterior probabilities of 78.9% for *A. lumbricoides*, 98.7% for *T. trichiura*, and 58.2% for hookworm infection. Model-based parameter estimates are summarised in Table 4.3. Temporal analysis shows that risks of infection with each soil-transmitted helminth have substantially reduced since 2005 (odds ratio estimated to be 0.47 for *A. lumbricoides*, 0.54 for *T. trichiura*, and 0.58 for hookworm infection).

We noted a negative association between *A. lumbricoides* infection and temperature variation measured by diurnal range and seasonality, and with high rainfall during the wettest season. Rainfall during the coldest quarter and acid soil with a pH between 5.4 and 5.8 were associated with an increased risk of infection. High temperatures during the warmest month and precipitation during the coldest quarter are associated with a higher risk of *T. trichiura* infection. The higher the altitude and the temperature variation, the lower the infection risk. Furthermore, a negative association was estimated with rainfall during the driest quarter (Table 4.3). With respect to hookworm infection, the risk increased with temperature. This risk is supported by a positive coefficient of minimum temperature during the coldest month and of the yearly temperature range after adjusting for the minimum temperature effect. Although precipitation during the wettest quarter increased the risk of hookworm infection, high soil moisture had a negative effect on the

Table 4.1: Overview of soil-transmitted helminth infection prevalence data in South America, by species and country.

	Identified reports*	Relevant papers	Survey location		Study level		Period	Survey date		Mean prevalence
			Total	Unique	Municipality	Point		Before 2005	2005 onwards	
A. lumbricoides	4022	159	2510	1682	2228	282	1952–2011	1456	1054	17.0%
Argentina	607	13	19	19	..	19	1993–2009	9	10	10.4%
Bolivia	70	7	49	49	7	42	1984–2006	28	21	27.6%
Brazil	1591	62	2219	1392	2180	39	2000–2009	1228	991	16.0%
Chile	554	14	28	28	2	26	1952–2007	27	1	14.2%
Colombia	133	18	19	19	1	18	1981–2008	10	9	17.1%
Ecuador	189	4	7	7	..	7	1989–2001	7	..	18.9%
French Guiana	137†	1	2	2	..	2	2000	2	..	14.4%
Guyana	137†	1	1	1	..	1	2002	1	..	18.8%
Paraguay	49	2	3	3	..	3	1996	3	..	4.3%
Peru	215	17	25	25	2	23	1986–2011	10	15	29.4%
Suriname	53	3	22	22	4	18	1963	22	..	30.5%
Uruguay	158	0	NA	NA	NA	NA	NA	NA	NA	NA
Venezuela	266	17	116	115	32	84	1984–2011	109	7	29.7%
T. trichiura	4022	137	2175	1506	1928	248	1952–2011	1252	924	9.3%
Argentina	607	12	18	18	..	18	1993–2009	7	11	3.9%
Bolivia	70	7	49	49	7	42	1984–2006	28	21	16.6%
Brazil	1591	51	1909	1241	1881	28	1968–2011	1045	864	7.4%
Chile	554	14	19	19	2	17	1952–2007	18	1	22.9%
Colombia	133	14	14	14	1	13	1981–2008	8	6	19.6%
Ecuador	189	3	5	5	..	5	1989–2001	5	..	9.0%
French Guiana	137†	0	NA	NA	NA	NA	NA	NA	NA	NA
Guyana	137†	1	1	1	..	1	2002	1	..	14.1%
Paraguay	49	0	NA	NA	NA	NA	NA	NA	NA	NA
Peru	215	16	24	23	1	23	1986–2011	10	14	22.4%
Suriname	53	3	22	22	4	18	1963	22	..	10.2%
Uruguay	158	0	NA	NA	NA	NA	NA	NA	NA	NA
Venezuela	266	16	115	114	32	83	1984–2011	108	7	31.8%
Hookworm	4022	101	2263	1541	2048	215	1963–2011	1331	932	9.4%
Argentina	607	9	15	15	..	15	1968–2009	7	8	38.6%
Bolivia	70	6	47	47	7	40	1993–2006	26	21	15.8%
Brazil	1591	48	2037	1315	2008	29	2000–2009	1151	886	8.8%
Chile	554	0	NA	NA	NA	NA	NA	NA	NA	NA
Colombia	133	10	11	11	..	10	1981–2008	7	4	4.3%
Ecuador	189	3	5	5	..	5	1989–2001	5	..	10.8%
French Guiana	137†	1	2	2	..	2	2000	2	..	63.0%
Guyana	137†	1	1	1	..	1	2002	1	..	28.2%
Paraguay	49	2	3	3	..	3	1963	3	..	21.3%
Peru	215	9	18	18	..	18	1991–2011	7	11	11.6%
Suriname	53	3	21	21	4	17	1963	21	..	34.3%
Uruguay	158	0	NA	NA	NA	NA	NA	NA	NA	NA
Venezuela	266	9	103	103	28	75	1984–2011	101	2	8.2%

Data are number, timeframe, or %.

.. = data not found.

NA = not applicable because there were no relevant papers for this country.

* Number of identified reports per country, irrespective of species.

† Number of identified reports of surveys done in either French Guiana or Guyana, or both.

Table 4.2: Variables Selected by a Gibbs variable selection procedure.

	***A. lumbricoides* infection**	***T. trichiura* infection**	**Hookworm infection**
Group 1*			
Yearly mean temperature	..	..	..
Maximum temperature of warmest month	..	Selected	..
Minimum temperature of coldest month	..	..	Selected
Mean temperature of wettest quarter	..	..	..
Mean temperature of driest quarter	..	..	..
Mean temperature of warmest quarter	..	..	..
Mean temperature of coldest quarter	..	..	..
Group 2*			
Mean diurnal temperature range	Selected	..	..
Yearly temperature range	..	Selected	Selected
Group 3*	..	..	..
Isothermality	..	..	..
Temperature seasonality	Selected	Selected	..
Group 4*			
Yearly precipitation	..	..	..
Precipitation in wettest month	Selected	..	..
Precipitation in wettest quarter	..	..	Selected
Group 5*			
Precipitation in driest month	..	..	..
Precipitation in driest quarter	..	Selected	..
Variables moderately correlated			
Precipitation seasonality	Selected	..	..
Precipitation in warmest quarter	..	..	..
Precipitation in coldest quarter	Selected	Selected	..
Altitude	..	Selected	..
Soil pH	Selected	..	..
Soil moisture	..	..	Selected
Posterior probability	78.9%	98.7%	58.2%

* Variables are grouped with other variables with which they are highly correlated.
A maximum of one variable per species per variable group was allowed to be selected.

disease.

Figures 4.2, 4.3 and 4.4 show model-based predictions of the geographical distribution of soil-transmitted helminth species-specific risk before and after 2005, and maps of the corresponding coefficient of variation of the predictive distribution and of the raw data. The three soil-transmitted helminth infections show spatial patterns that are affected by the selected environmental predictors. Spatial correlations extended to distances ranging

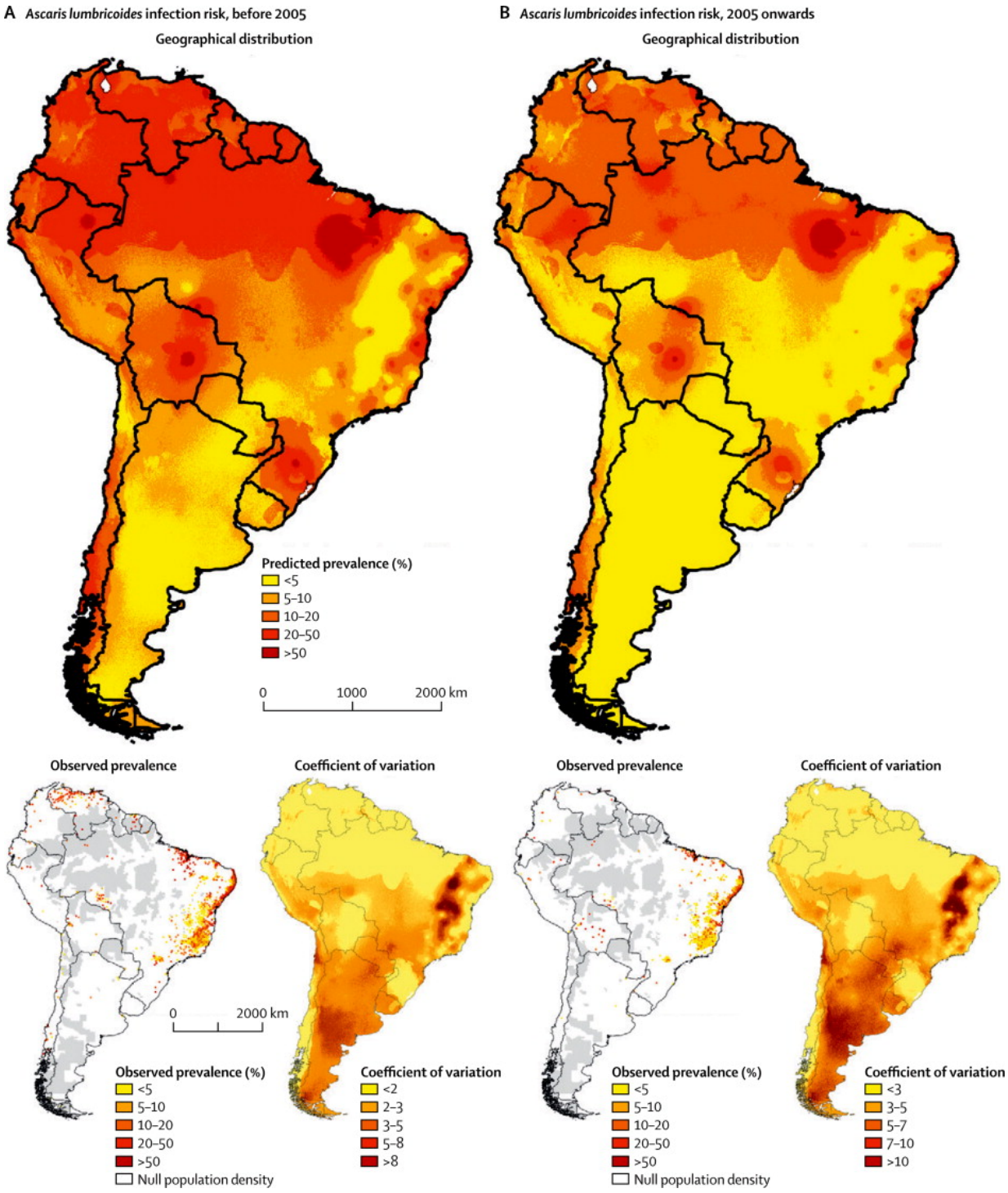


Figure 4.2: *Ascaris lumbricoides* infection risk in South America. Infection risk estimated from surveys done before 2005 (A) and from 2005 onwards (B).

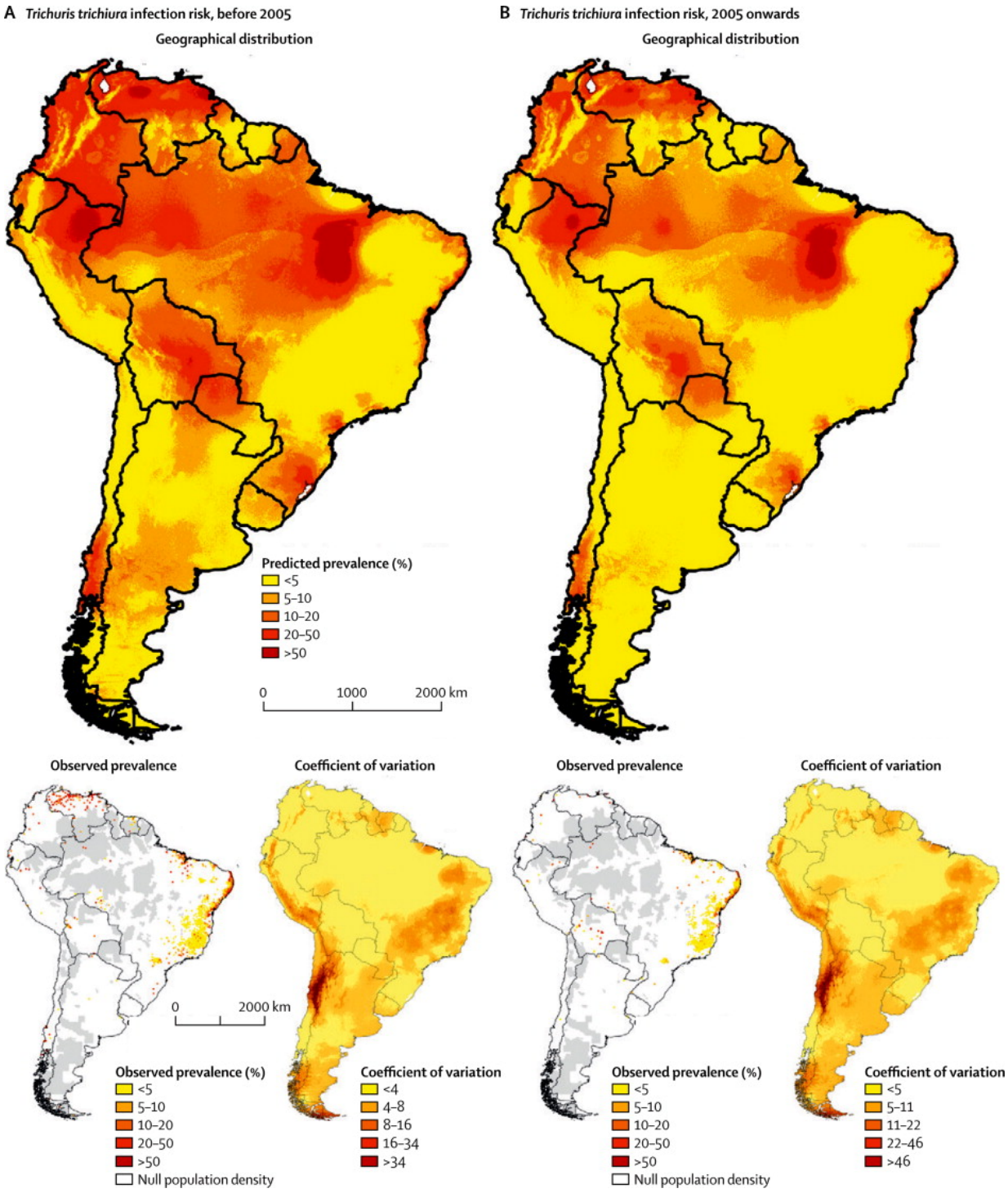


Figure 4.3: *Trichuris trichiura* infection risk in South America. Infection risk estimated from surveys done before 2005 (A) and from 2005 onwards (B).

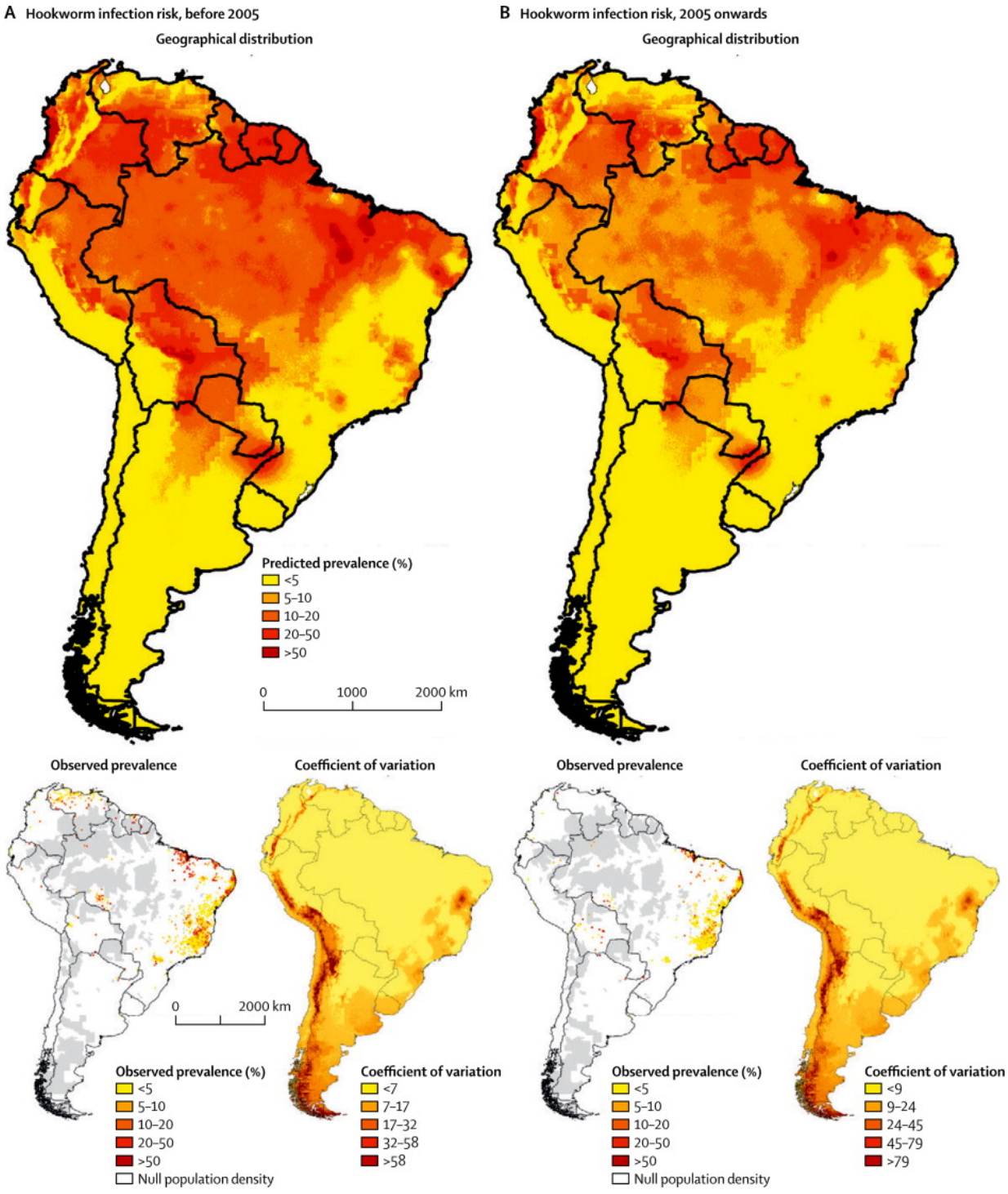


Figure 4.4: Hookworm infection risk in South America. Infection risk estimated from surveys done before 2005 (A) and from 2005 onwards (B).

between 465.7 km and 905.3 km depending on the helminth species. The north of the subcontinent had a higher risk of infection compared with the south. Prediction uncertainty was more important for regions with sparse data, as shown in the maps of the coefficient of variation of the predictive distribution.

Population-adjusted prevalence estimates at country and subcontinental level for 2005 onwards are presented in Table 4.4. The estimated *A. lumbricoides* infection risk (15.6%) was higher than the risk of *T. trichiura* (12.5%) and hookworm (11.9%) infection. On the assumption that risk of infection with the three soil-transmitted helminths was independent, we predicted an overall risk of infection with a soil-transmitted helminth of 27.1%. The countries at highest risk are French Guiana for *A. lumbricoides* infection (25.1%) and hookworm infection (36.6%), and Venezuela for *T. trichiura* infection (27.3%). French Guiana has the highest predicted risk for any soil-transmitted helminth (46.2%).

Model validation results are given in Figure 4.5. The risks of infection are correctly predicted within a 95% CI for 98.6% of locations for *A. lumbricoides*, 96.5% for *T. trichiura*, and 97.8% for hookworm. The mean error based on the median of the posterior distribution was 0.3%, −3.4%, and −2.5% for the three species, suggesting that our models slightly underestimate the prevalence of *T. trichiura* and hookworm infection. Table 4.5 presents the population at high (prevalence $>$ 50%) and low (prevalence 20–50%) infection risk, stratified by country and soil-transmitted helminth species. The numbers of people living in regions with risk greater than 20% are estimated to be 18.7 million for *A. lumbricoides*, 30.4 million for *T. trichiura*, and 13.3 million for hookworm infection.

4.4 Discussion

We present the first model-based estimates of soil-transmitted helminth infection risk and number of infected people at a detailed geographical scale for South America. Across the subcontinent, *A. lumbricoides* infection was most prevalent (61.5 million people were estimated to be infected; 15.6%), followed by *T. trichiura* (49.3 million; 12.5%) and hookworm (46.9 million; 11.9%).

Mechanisms of interaction between the three soil-transmitted helminth species have yet to be fully elucidated. Co-infection with several helminths might impair the host's immune response to single parasites, and might increase susceptibility to clinical disease. Co-infection is probably high because the three helminths share similar climatic and socioeconomic ranges (Brooker et al., 2006; Pullan and Brooker, 2012). Although interspecies associations in hosts with several infections have been suggested (Booth et al., 1998), there

is no evidence to show that the three infections are heavily dependent. In our analysis, we assumed that at an individual level, the chances of being infected with any one helminth species were the same, irrespective of the presence of another species, and we estimated that the overall soil-transmitted helminthiasis risk is 27.1%.

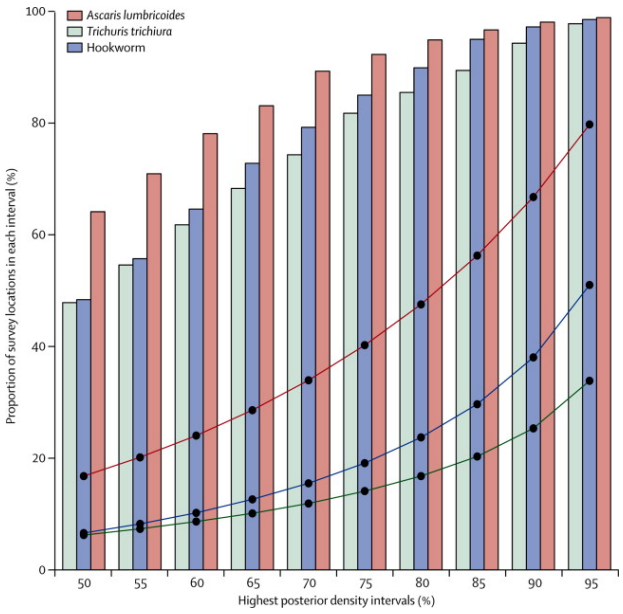


Figure 4.5: Model validation results. Proportion of survey locations with prevalence of infection falling in the predicted highest posterior density intervales (bar plots) for *Ascaris lumbricoides*, *Trichuris trichiura*, and hookworm. The line plots show the corresponding width of the predicted highest density region.

Table 4.3: Logistic regression parameter estimates for *Ascaris lumbricoides*, *Trichuris trichiura*, and hookworm infection risk.

	Bivariate, non-spatial, odds ratio (95% CI)	**Multivariable geostatistical odds ratio (95%BCI)**
***A. lumbricoides* infection**		
Period		
< 2005	1.00	1.00
≥ 2005	0.52 (0.52; 0.53)*	0.47 (0.46; 0.47)*
Mean diurnal temperature range	0.56 (0.56; 0.56)*	0.81 (0.77; 0.90)*
Temperature seasonality	0.52 (0.52; 0.52)*	0.92 (0.85; 0.99)*
Precipitation in wettest month	1.35 (1.34; 1.35)*	0.97 (0.93; 0.99)*
Precipitation seasonality	0.87 (0.87; 0.87)*	1.18 (1.11; 1.25)*
Soil pH		
< 5.4	1.00	1.00
5.4–5.8	0.73 (0.72; 0.73)*	1.14 (1.07; 1.17)*
≥ 5.8	0.80 (0.80; 0.80)*	0.93 (0.83; 1.02)
Precipitation in coldest quarter (mm)		
< 80	1.00	1.00
80–320	1.47 (1.47; 1.48)*	1.12 (0.99; 1.30)
≥ 320	4.04 (4.03; 4.06)*	2.00 (1.86; 2.31)*
τ^2 (non-spatial variance)		1.49 (1.37; 1.68)
σ^2 (spatial variance)		2.59 (1.85; 3.77)
Range (km)		507.8 (336.6; 807.7)
***T. trichiura* infection**		
Period		
< 2005	1.00	1
≥ 2005	0.59 (0.59; 0.59)*	0.54 (0.54; 0.55)*
Maximum temperature of warmest month	1.23 (1.22; 1.23)*	1.40 (1.21; 1.47)*
Precipitation in coldest quarter (mm)		
< 80	1.00	1.00
80–335	2.24 (2.23; 2.26)*	0.96 (0.91; 1.02)
≥ 335	6.84 (6.80; 6.88)*	1.34 (1.25; 1.39)*
Altitude (m)		
< 200	1.00	1.00
200–600	0.36 (0.36; 0.36)*	0.80 (0.76; 0.87)*
≥ 600	0.12 (0.12; 0.12)*	0.70 (0.62; 0.73)*
Yearly temperature range (°C)		
< 13.0	1.00	1.00
13.0–17.5	0.26 (0.26; 0.26)*	0.53 (0.51; 0.57)*
≥ 17.5	0.13 (0.13; 0.14)*	0.66 (0.60; 0.74)*

Continued on next page

	Bivariate, non-spatial, odds ratio (95% CI)	**Multivariable geostatistical odds ratio (95%BCI)**
Temperature seasonality (°C)		
< 1.25	1.00	1.00
1.25–1.65	0.65 (0.65; 0.66)*	0.84 (0.81; 0.89)*
≥ 1.65	0.19 (0.19; 0.19)*	0.79 (0.73; 0.83)*
Precipitation in driest quarter (mm)		
< 50	1.00	1.00
50–100	1.19 (1.19; 1.20)*	0.92 (0.85; 0.99)*
≥ 100	3.23 (3.22; 3.25)*	0.81 (0.72; 0.88)*
τ^2 (non-spatial variance)		1.44 (1.31; 1.61)
σ^2 (spatial variance)		4.21 (2.88; 6.11)
Range (km)		905.3 (591.4; 1451.1)
Hookworm infection		
Period		
< 2005	1.00	1.00
≥ 2005	0.62 (0.62; 0.63)*	0.58 (0.58; 0.59)*
Minimum temperature of coldest month	2.02 (2.01; 2.02)*	2.13 (1.99; 2.18)*
Yearly temperature range	0.64 (0.64; 0.64)*	1.35 (1.33; 1.41)*
Precipitation in wettest quarter	1.47 (1.47; 1.48)*	1.55 (1.44; 1.69)*
Soil moisture (%)		
< 50	1.00	1.00
50–80	0.99 (0.99; 1.00)	0.59 (0.56; 0.60)*
≥ 80	0.65 (0.64; 0.65)*	0.31 (0.26; 0.41)*
τ^2 (non-spatial variance)		2.09 (1.86; 2.30)
σ^2 (spatial variance)		2.68 (1.67; 3.66)
Range (km)		465.7 (292.8; 683.1)

* Calculated under the assumption that *A. lumbricoides*, *T. trichiura*, and hookworm infections are independnet of one another.

Data are population-adjusted prevalence (95% Bayesain credible interval) based on 2010 population estimates.

Table 4.4: Population-adjusted prevalence of helminth infection from 2005 onwards, by country.

	Ascaris lumbricoides	***Trichuris trichiura***	**Hookworm**	**All soil-transmitted helminth infections***
South America	15.6% (14.4; 17.2)	12.5% (11.1; 13.6)	11.9% (10.5; 13.4)	27.1% (25.9; 28.4)
Argentina	9.5% (6.9; 13.8)	8.5% (6.3; 11.6)	8.3% (6.6; 10.3)	18.9% (16.4; 23.0)
Bolivia	24.5% (18.4; 29.8)	18.9% (14.5; 24.9)	19.3% (15.2; 23.5)	37.8% (33.2; 41.6)
Brazil	14.3% (13.4; 16.2)	10.1% (8.8; 11.3)	12.3% (10.8; 14.1)	25.5% (24.2; 27.1)
Chile	16.8% (13.2; 19.9)	11.7% (8.4; 15.4)	7.2% (5.5; 8.8)	24.7% (21.5; 27.9)
Colombia	19.7% (16.9; 24.0)	17.8% (14.4; 21.7)	14.5% (10.3; 17.5)	33.4% (30.7; 37.6)
Ecuador	18.5% (14.9; 24.9)	11.1% (8.4; 15.1)	12.3% (8.1; 15.7)	28.1% (23.8; 33.3)
French Guiana	25.1% (15.3; 33.0)	15.1% (7.4; 29.9)	36.6% (28.8; 44.3)	46.2% (38.4; 56.1)
Guyana	23.6% (16.4; 31.6)	11.1% (5.3; 16.7)	25.0% (19.5; 29.6)	37.4% (31.9; 42.9)
Paraguay	8.3% (5.3; 15.5)	7.1% (2.6; 21.4)	18.3% (12.1; 28.0)	26.7% (20.1; 34.7)
Peru	17.3% (13.4; 20.6)	10.8% (8.4; 13.6)	8.7% (6.1; 11.3)	25.0% (21.3; 29.0)
Suriname	23.5% (15.9; 34.8)	8.3% (3.5; 16.6)	30.0% (20.3; 41.3)	40.1% (30.2; 51.1)
Uruguay	9.9% (6.9; 16.1)	10.1% (7.3; 15.1)	6.4% (4.2; 9.2)	18.8% (14.7; 24.7)
Venezuela	21.7% (17.9; 25.4)	27.3% (25.0; 29.8)	12.2% (9.8; 15.6)	39.4% (36.3; 42.5)

* Calculated under the assumption that *A. lumbricoides*, *T. trichiura*, and hookworm infections are independnet of one another.

Data are population-adjusted prevalence (95% Bayesian credible interval) based on 2010 population estimates.

Table 4.5: Estimated population at risk of soil-transmitted helminth infection, by country.

	Population at low (20–50%) risk (x1000)[†]			**Population at high (≥ 50%) risk (x1000)**[†]			**Total population (x1000)**[*]
	Ascaris lumbricoides	***Trichuris trichiura***	**Hookworm**	***Ascaris lumbricoides***	***Trichuris trichiura***	**Hookworm**	
South America	15,923	28,190	11,702	2,752	2,169	1,572	394,234
Argentina	0	0	1,101	0	0	16	47,417
Bolivia	1,510	2,030	2,201	927	0	594	10,453
Brazil	10,681	8,963	5,853	1,782	1,597	290	186,552
Chile	258	345	0	0	0	0	19,435
Colombia	1,131	4119	1258	0	0	671	48,235
Ecuador	26	3	249	0	0	0	14,118
French Guiana	0	0	125	0	0	0	145
Guyana	1	0	79	0	0	0	629
Paraguay	0	0	513	0	0	0	7,247
Peru	828	347	62	42	547	1	27,832
Suriname	0	0	51	0	0	0	407
Uruguay	0	0	0	0	0	0	3,977
Venezuela	1,487	12,382	209	0	26	0	27,786

[*] Estimates based on 2010 data from Gridded Population of the World, version 3 (CIESIN, FAO and CIAT, 2005).

[†] Calculations were based on the median of the posterior distribution of the predicted risk from 2005 onwards.

Our geostatistical meta-analysis estimated transmission dynamics and broad-scale geographical patterns of soil-transmitted helminthiasis risk at continental and national levels in South America. However, our model-based, high spatial resolution estimates should be interpreted cautiously, because our systematic review showed that data was scarce and that national surveys are needed to obtain up-to-date information about infection risk. Brazil is the only country with a large number of soil-transmitted helminth surveys, which is partly explained by the establishment of a national schistosomiasis control programme in 1986 (Minsiterio da Saúde-Brasil, a). As part of monitoring, evaluation, and surveillance of this programme, which repeatedly assesses the prevalence of *Schistosoma mansoni* with the Kato-Katz technique, stool samples were concurrently examined for soil-transmitted helminths (Minsiterio da Saúde-Brasil, b). Indeed, areas endemic for schistosomiasis had many survey locations, whereas other parts of the country, especially the Amazon rainforest, had only very little information. No data were available for Uruguay, and we identified very few surveys for Argentina, Chile, Ecuador, French Guiana, Guyana, and Paraguay. Because of data scarcity, modelling risk at the unit of the country would not have been feasible. Our aim was to produce a map for the whole subcontinent and to obtain estimates in countries with little or no information. In regions with sparse or no data, our predictions are based on data available in neighbouring countries and on climatic factors identified to drive their spatial distribution. In the absence of control interventions, the assumption that infection risk is governed by climatic suitability is reasonable. Therefore, we believe that our predictions provide valuable information, although the results should be interpreted with caution, particularly in regions where data are sparse. New surveys should be done in southern countries such as Argentina and Uruguay to validate our model-based soil-transmitted helminth risk predictions of below 20%. However, we propose that first priority should be given to surveys in countries such as French Guiana, Suriname, Venezuela, Bolivia, Guyana, Colombia, Ecuador, Paraguay, Peru, and Chile, for which our models predict that there are regions where prevalences are above 20%, and therefore need interventions according to PAHO recommendations (PAHO and WHO, 2009). A soil-transmitted helminth prevalence survey has recently been done in five districts in Suriname. These data should be included in subsequent geostatistical analyses once they become available.

The Bayesian variable selection approach used in our analysis identified important predictors related to the biology, ecology, and epidemiology of soil-transmitted helminthiasis.

The effect of some covariates changed after accounting for other covariates or spatial correlation. This issue emphasises the need to account for spatial correlation during analysis of geographical data. Warm and humid climatic conditions have been reported to be suitable for the development of parasite eggs and larvae (Spindler, 1929; Brooker et al., 2006). Our analyses confirm the positive association of *T. trichiura* and hookworm infection risks with proxies related to temperature and precipitation, and of *A. lumbricoides* with rainfall. Results of earlier studies provided temperature limits for the survival of the larvae of each soil-transmitted helminth species, and suggested that extreme temperatures might slow down or even stop their development (Brooker et al., 2004b). This effect is evident in the *A. lumbricoides* and *T. trichiura* models, which estimated a negative association between temperature variation and infection risks. The negative effect of altitude on *T. trichiura* infection might arise as a result of lower temperature at higher altitudes. *A. lumbricoides* infection is more prevalent in regions with acid soil. Similar results had been reported by Yadav (2003), who considered different levels of soil acidity and reported that *A. lumbricoides* egg development was accelerated in soils with a pH of 5.0 under specific moisture and temperature conditions. Our study estimated a negative effect of moist soil on hookworm risk after adjusting for precipitation during the wettest quarter. We also estimated a negative effect of rainfall on the risk of *T. trichiura* during the driest quarter after adjusting for precipitation during the coldest quarter. These observations support earlier interpretations that high precipitation might wash out the helminth eggs and therefore reduce infection (Brown, 1927; Gunawardena et al., 2004).

Our maps of the geographical distribution of soil-transmitted helminth infection risks present patterns coherent with recently released maps of the transmission limits of these infections (Pullan and Brooker, 2012). More precisely, the estimated risk for any of the three soil-transmitted helminths is negligible in the northern part of Chile. The risk of infection with *A. lumbricoides* and hookworm is low in Argentina. Hookworm risk is minor in the coastal regions of Peru and southern Chile. Although several regions were within the climatic suitability transmission limits provided by Pullan and Brooker (2012), our models have identified additional regions where the estimated risk is of little importance ($< 5\%$, population-unadjusted) — i. e. Uruguay, Paraguay, and parts of Brazil for *A. lumbricoides*; Guyana, Suriname, and parts of Brazil for *T. trichiura*; and coastal regions of Venezuela and parts of Brazil for hookworm. In South America, the population living in regions with a predicted risk greater than 20% is higher for *T. trichiura* (30.4 million) than for *A. lumbricoides* (18.7 million) and hookworm (13.3 million), suggesting that *T. trichiura*

has a particularly high risk in densely populated regions. This result is in agreement with previous findings of a greater *T. trichiura* risk in peri-urban, rather than rural, regions (Pullan and Brooker, 2012).

The surveys used different diagnostic approaches varying in sensitivity and specificity (Goodman et al., 2007; Knopp et al., 2008; Bergquist et al., 2009), which might have introduced bias into the data. Importantly though, the WHO-recommended Kato-Katz technique (Montresor et al., 1998) was used in more than 95% of the surveys. Because of data sparsity, we did not exclude the remaining 5% of surveys, which used various techniques (Appendix 4.5.6). Of the studies using the Kato-Katz technique, we noted that sampling efforts differed between surveys. Although most surveys prepared two slides, some used one Kato-Katz thick smear, or combined the Kato-Katz technique with other diagnostic techniques. Differences in sensitivity and specificity can be incorporated into our Bayesian models (Wang et al., 2008). However, more than 10% of the surveys did not provide detailed information about diagnostic methods (including sampling efforts). Hence, any related assumption would be debatable and introduce additional bias. Additionally, assessment of prior information about the various sampling efforts would have been difficult.

Survey design also differs. Children aged 5–14 years are known to carry the heaviest burden of soil-transmitted helminthiasis (Anderson and May, 1985), but 98% of our survey locations were in the community. Of the remaining 2%, the surveys either screened children or did not provide any details of the population examined. We included surveys in the meta-analysis irrespective of the age profile, which might be sources of bias. Geostatistical models with age-alignment factors (Schur et al., 2011a; Soares Magalhães and Clements, 2011) have been developed by calculating country-specific alignment factors. In accordance with this approach, some countries would have had insufficient data to estimate the alignment factor. Furthermore, adjusting for age would have required some arbitrary assumptions about the unknown population and might have introduced even more bias.

We did not consider socioeconomic proxies when building our models. Because soil-transmitted helminth infections disproportionally affect poor people, such proxies might be related to infection risk. For some countries, measures of the proportion of households with sanitation are available from census data, but they often contain missing data and their spatial resolution is low. However, in previous studies, these poverty proxies did not explain small-scale spatial variation when diseases were modelled over large scales, which is intrinsically linked to the issue of pattern and scale (Levin, 1992). More efforts are needed to construct a socioeconomic index or to compile relevant socioeconomic predictors

that can explain the socioeconomic disparities at high spatial resolution. A temporal trend was estimated by a binary covariate indicating whether surveys had been done before or after 2005. We could not take into account spatio-temporal correlations because of the coarse temporal resolution of the data. Our analysis suggests that the risk of soil-transmitted helminth infection has decreased substantially in recent years in South America (De Silva et al., 2003). We believe that this decrease is more related to the improvement of socioeconomic factors and on-going control measures during the past decade than to climate change. *A. lumbricoides* risk seems to have reduced more than that of the other soil-transmitted helminth infections. This risk reduction might be related to the higher efficacy of the available treatments against *A. lumbricoides* (Keiser and Utzinger, 2008), leading to fewer eggs released into the environment and, consequently, a lower risk of reinfection.

South America has high climatic heterogeneity. Hence, a single logistic regression equation might not be able to describe variation of the effect of climatic factors on the disease risk across the subcontinent. A model estimating the climate–disease risk relation by ecological zone could be considered. However, this model would need better coverage of the ecological zones, with new survey data. Furthermore, we assumed a single spatial process over the whole of South America, which might not be true because of the region's large climatic diversity, socioeconomic disparities, and health system inequalities (Belizán et al., 2007). Non-stationary models assuming a mixture of separate stationary processes within each agroecological zone have been applied to schistosomiasis and malaria risk mapping (Raso et al., 2006a; Gosoniu et al., 2009). A similar approach could be implemented and applied to a large dataset by approximating each spatial process on a subsample of locations.

Another limitation to consider is that people might have been infected before the survey period began. Even if seasonal dynamics in transmission occur, their overall effect on transmission dynamics is limited because of the lifespan of adult worms, which exceeds 1 year (Anderson and May, 1985). Therefore, we did not include interactions of climatic factors with seasons.

In conclusion, our study provides open-access compilation of available soil-transmitted helminth infection prevalence survey data, gives model-based georeferenced estimates of infection risk, and identifies regions with sparse or no data across South America. We believe that our estimates provide a valuable assessment of the national and subnational situation of soil-transmitted helminthiasis, and we hope that this first effort will contribute

useful information to national control programmes for planning control and elimination strategies by indicating regions where survey data are urgently needed if risk prediction is to be improved. We will continue our efforts to further develop statistical models and update the GNTD database. We also invite researchers to assist these efforts by contributing their georeferenced data to the GNTD database so it can be further developed into a useful shared platform for research and control of soil-transmitted helminthiasis and other neglected tropical diseases.

Acknowledgements

We thank the Pan American Health Organization Neglected Infectious Diseases Programme for their contribution to the data collection. The authors are grateful for financial support from the UBS Optimus Foundation and the Brazilian Swiss Joint Research Programme (BSJRP 011008).

4.5 Appendix

4.5.1 Soil-transmitted helminthiasis systematic review protocol

Literature search
- Peer-reviewed via PubMed (2012) and ISI Web of Knowledge (2012) with following keywords: argentin* (OR bolivi*, OR brazil, OR brasil, OR chile, OR colombi*, OR ecuad*, OR guyan*, OR guian*, OR paragu*, OR peru, OR surinam*, OR urugu* OR venezuel*) AND helminth* (OR ascari*, OR trichur*, OR hookworm*, OR necator, OR ankylostom*, OR ancylostom*,OR strongy*, OR hymenolepis, OR toxocara, OR enterobius*, OR geohelminth*, OR nematode*)
- Unpublished grey literature (Ministry of Health report and report, personal communication)

↓

Identification of potentially relevant publications. Review of titles and abstracts, if available, to identify potentially relevant publications. Main exclusion criteria are: publications pertaining on animals/plants/genetic, case reports, in-vitro studies, and/or absence of soil-transmitted helminthiasis surveys.

↓

First quality check. Random selection of 30% of the papers identified as irrelevant for quality assessment of relevance (this percentage is adapted to the experience of the personnel which carried out the search). In case of any misclassification, the entire country is double-checked.

↓

Review of the potentially relevant publications. Full review of papers identified as potentially relevant. Main exclusion criteria are: absence of prevalence data, specific patient group (hospital-based, HIV, newborns...), case-control, clinical trials or pharmacological studies (except control group), displaced population (travellers, military...), population dewormed during the past year.

↓

Geolocation. Retrospective geolocation of study locations using information provided in the publications, if any, and various online sources (Wikimapia, Google Maps, iGuide Interactive Travel Guide...). Centroid of the administrative unit is assigned to areal data, calculated from the administrative boundaries maps of the Database of Global Administrative Areas (GADM v1).

↓

Prevalence data extraction. Relevant prevalence data are entered in the GNTD database with (i) the source: authors, journal, publication date, (ii) the description of the survey: date, type of survey, (iii) the location information: coordinates, name, administrative unit, (iv) the parasitological data: species, number of positive/examined, prevalence, age, diagnostic tool.

↓

Contact authors. In case of missing information, authors are contacted if the paper has been published in the past 20 years.

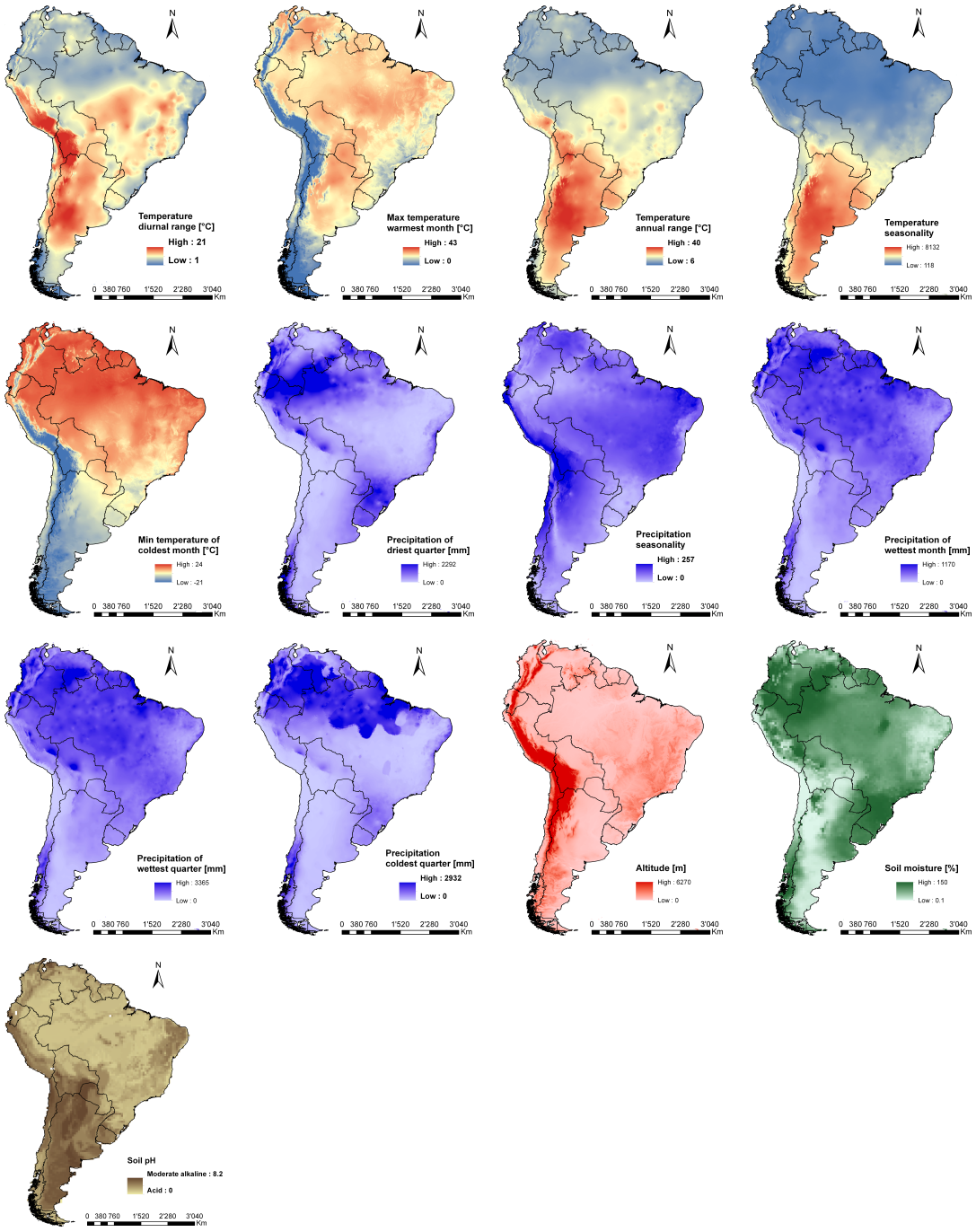
↓

Second quality check. All publications entered in the GNTD are assessed for quality of extraction, and all coordinates are double-checked in Google Maps for quality assessment of the geolocation.

↓

Meta-analysis. Survey where sample size < 10 are excluded from the analysis. Sample size $n = 100$ is assigned to surveys where only prevalence was given. Date of publication is assigned to the survey if the date of the survey is missing. Data are screened by location and if there is evidence of duplicate, the survey with the most complete information is kept for the analysis.

4.5.2 Spatial distribution of the climatic and environmental predictors in South America



4.5.3 Gibbs variable selection

Predictors highly correlated were grouped into K groups k , $k = (1, ..., K)$, each containing $J^{(k)}$ highly correlated variables $j^{(k)}$, $j^{(k)} = (1, ..., J^{(k)})$. We let Y_i and n_i be the number of infected and the number of screened individuals, and p_i be the probability of infection at location i, $i = (1, ..., N)$. Assuming that Y_i follows a binomial distribution $Y_i \sim Bn(n_i, p_i)$, the survey period T_i, the design matrix for the kth group for the $j^{(k)}$th predictors $Z_{ji}^{(k)}$ and a location specific random effect ε_i are modelled on the logit scale; $\log it(p_i) = \alpha_0 + \alpha_1 * \mathrm{T_i} + \sum_j \sum_k I_{j^{(k)}} \beta_{j^{(k)}}^T Z_{ji}^{(k)} + \varepsilon_i$.

For each predictor, we introduce a multinomial indicator with $J^{(k)} + 1$ categories, where $J^{(k)}+1$ indicates the absence of any of the predictors in the group k, such as $(I_1, ..., I_{J^{(k)}+1}) \sim Mult(1, 1/(J^{(k)} + 1), ..., 1/(J^{(k)} + 1))$, and we assume a mixture prior distribution for $\beta_{j^{(k)}}$ such that $\beta_{j^{(k)}} \sim I_{j^{(k)}} N(0, \sigma_k^2) + (1 - I_{j^{(k)}}) N(0, \tau_k^2)$, where σ_k^2 is the variance parameter set to be 100 when the variable $Z_j^{(k)}$ is included in the model, and τ_k^2 is the variance set to be 0.001 when the variable $Z_j^{(k)}$ is excluded from the model. The $\varepsilon_i' s$ are assumed to be Gaussian $\varepsilon_i \sim N(0, \tau^2)$.

Markov chain Monte Carlo (MCMC) simulations were used to estimate model parameters in WinBUGS 14 (Imperial College and Medical Research Council; London, UK). Models ran an average of 100000 iterations with one chain sampler. A sample of 20000 estimates of each model parameter was used to determine the model with the highest posterior probability.

4.5.4 Geostatistical model and model validation

The model described in Appendix 4.5.3 was extended to a geostatistical analogue by introducing spatially structured random effects φ_i at locations i. In particular, $\log it(p_i) = X_i^T \beta + \varepsilon_i + \varphi_i$, where X_i is the vector of covariates selected by the Gibbs variable selection procedure. We assume that φ_i's follow a latent stationary Gaussian process, $\varphi = (\varphi_1, ..., \varphi_N)^T \sim MVN(0, \Sigma)$ with variance-covariance matrix Σ, $\Sigma_{jk} = \sigma^2 \exp(-\rho d_{jk})$, where d_{jk} is the Euclidean distance between locations j and k, σ^2 is the geographical variability known as the partial sill and ρ is a smoothing parameter controlling the rate of correlation decay. The geographic dependence (range), is defined as the minimum distance at which spatial correlation between locations is less than 5% and is calculated by $3/\rho$.

Large matrix computations are overcome by estimating the spatial process from a subset of 200 locations (knots) $\{s_l^*, l = 1, .., L\}$ with latent observations $\varphi^* = (\varphi_1^*, ..., \varphi_L^*)^T$,

where $\varphi^* \sim MVN(0, \Sigma^*)$, Σ^* is the LXL covariance matrix approximating Σ. The spatial random effects φ at the original set of locations are predicted via the conditional mean $Q^T \Sigma^{*-1} \varphi^*$, where $Q = Cov(\varphi^*, \varphi)$ is a NXL matrix of the covariance functions between the L knots and the N observed locations. The knots of the subset were selected using the cover.design routine in R (R Core Team, 2014) which minimises a geometric space-filling criterion. Semivariogram analyses suggested that a sample of 200 knots was sufficient to preserve the spatial correlation in the data. Within the Bayesian framework of inference, non-informative uniform prior distributions were adopted for the regression coefficients β's. A non-informative gamma prior was chosen for ρ and vague inverse-gamma priors for the variances σ^2 and τ^2.

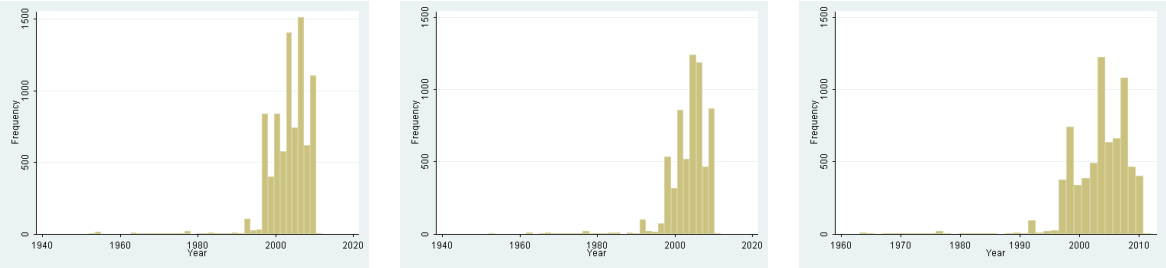
Markov chain Monte Carlo (MCMC) simulations were implemented in Fortran 95 (Compaq Visual Fortran Professional 6.6.0) code written by the authors using standard numerical libraries (NAG, The Numerical Algorithm Group Ltd.). All models were run with one chain samplers and a burn-in of 5000 iterations. Convergence was assessed by employing the Raftery and Lewis (1992) diagnostics. Bayesian kriging was performed using a posterior sample of size 500.

Predictive ability of the models were assessed by calculating the proportion of test locations correctly predicted within the kth highest posterior density interval with different probability coverages of the posterior distribution.

The mean error (ME) assessed bias of the prediction and was calculated by averaging the differences between the median of the posterior predictive distribution and the observed prevalence over all the test locations.

4.5.5 Period distribution of the prevalence survey data

A) *A. lumbricoides* B) *T. trichiura* C) Hookworm



4.5.6 Number and percentage of compiled surveys stratified by diagnostic techniques and parasite species

Diagnostic technique	***A. lumbricoides***	***T. trichiura***	**Hookworm**
Kato-Katz	2343 (93.3%)	2036 (93.6%)	2156 (95.3%)
Kato-Katz in combination with other techniques	15 (0.6%)	15 (0.7%)	11 (0.5%)
Sedimentation techniques[1]	62 (2.5%)	50 (2.3%)	44 (1.9%)
Concentration techniques[2]	25 (1.0%)	27 (1.2%)	16 (0.7%)
Direct smear	5 (0.2%)	3 (0.1%)	3 (0.1%)
Not stated	60 (2.4%)	44 (2.0%)	33 (1.5%)

[1] Include stool centrifugation, Ritchie and Baermann techniques.
[2] Include Teleman modified and Faust techniques.

4.5.7 Number and percentage of prevalence survey data included in the meta-analysis with missing information stratified by parasite species

Missing information	***A. lumbricoides***	***T. trichiura***	**Hookworm**
Survey period	98 (3.9%)	86 (4.0%)	69 (3.0%)
Number of individuals screened	7 (0.3%)	7 (0.3%)	7 (0.3%)
Age population screened	18 (0.7%)	18 (0.8%)	18 (0.8%)
Kato-Katz test			
without number of stool specimen	131 (5.6%)	141 (6.9%)	123 (5.7%)
without number of thick smear examined	248 (10.5%)	243 (11.8%)	226 (10.4%)

Chapter 5

Modelling the geographical distribution of soil-transmitted helminth infections in Bolivia

Chammartin F.[1,2], Scholte R.G.C.[1,2], Malone J.B.[3], Bavia M.E.[4], Nieto P.[3], Utzinger J.[1,2] and Vounatsou P.[1,2]

[1] Swiss Tropical and Public Health Institute, Basel, Switzerland
[2] University of Basel, Basel, Switzerland
[3] School of Veterinary Medicine, Louisiana State University, Baton Rouge, LA, United States of America
[4] Preventive Medicine Department, Federal University of Bahia, Salvador, Bahia, Brazil

This paper has been published in *Parasites & Vectors* 2013, 6: 152.

Abstract

Background: The prevalence of infection with the three common soil-transmitted helminths (i. e. *Ascaris lumbricoides*, *Trichuris trichiura*, and hookworm) in Bolivia is among the highest in Latin America. However, the spatial distribution and burden of soil-transmitted helminthiasis are poorly documented.

Methods: We analysed historical survey data using Bayesian geostatistical models to identify determinants of the distribution of soil-transmitted helminth infections, predict the geographical distribution of infection risk, and assess treatment needs and costs in the frame of preventive chemotherapy. Rigorous geostatistical variable selection identified the most important predictors of *A. lumbricoides*, *T. trichiura*, and hookworm transmission.

Results: Results show that precipitation during the wettest quarter above 400 mm favours the distribution of *A. lumbricoides*. Altitude has a negative effect on *T. trichiura*. Hookworm is sensitive to temperature during the coldest month. We estimate that 38.0%, 19.3%, and 11.4% of the Bolivian population is infected with *A. lumbricoides*, *T. trichiura*, and hookworm, respectively. Assuming independence of the three infections, 48.4% of the population is infected with any soil-transmitted helminth. Empirical-based estimates, according to treatment recommendations by the World Health Organization, suggest a total of 2.9 million annualised treatments for the control of soil-transmitted helminthiasis in Bolivia.

Conclusion: We provide estimates of soil-transmitted helminth infections in Bolivia based on high-resolution spatial prediction and an innovative variable selection approach. However, the scarcity of the data suggests that a national survey is required for more accurate mapping that will govern spatial targeting of soil-transmitted helminthiasis control.

5.1 Background

Soil-transmitted helminth infections are mainly caused by the intestinal worms *Ascaris lumbricoides*, *Trichuris trichiura*, and the two hookworm species *Ancylostoma duodenale* and *Necator americanus* (Bethony et al., 2006). They are the most prevalent neglected tropical diseases, and they are widely distributed across Latin America (Schneider et al., 2011; Pullan and Brooker, 2012). Soil-transmitted helminthiasis and other neglected tropical diseases primarily affect low-income populations, causing chronic conditions, learning disabilities, and reduced productivity and income earning capacity in later life. Morbidity control and, where resources allow, local elimination are now recognised as a priority for achieving the millennium development goals (Hotez et al., 2008b). In 2009, the Pan American Health Organization (PAHO) developed a plan to eliminate neglected and other poverty-related diseases in Latin America and Caribbean countries. Soil-transmitted helminthiases were identified as target diseases to be controlled through preventive chemotherapy and by promoting access to clean water, improved sanitation, and better hygiene behaviour (PAHO and WHO, 2009). Control programmes require reliable baseline information of the geographical distribution of the number of infected people and disease burden estimates in order to enhance the spatial targeting and cost-effectiveness of planned interventions (Simoonga et al., 2009; Soares Magalhães et al., 2011).

Bolivia is ranked last among the Western Hemisphere countries in terms of key health indicators. For example, child mortality rate is the worse in South America and, according to the 2001 census, 64% of the population did not have enough income to meet their basic needs. The prevalence of soil-transmitted helminth infection is estimated at around 35% (IADB, PAHO and SABIN, 2011). However, the geographical distribution and burden of soil-transmitted helminth infections is poorly documented.

In the past 20 years, progress in geographical information system (GIS) and remote sensing techniques, coupled with spatial modelling, enabled a better understanding of helminth ecology and mapping at high spatial resolution (Brooker et al., 2002; Raso et al., 2005; Clements et al., 2009b; Simoonga et al., 2009; Soares Magalhães et al., 2011; Schur et al., 2011a). Ecological niche and biology-driven models have been used in assessing the distribution of helminth infections (Malone, 2005; Mudenda et al., 2012; Stensgaard et al., 2013). Bayesian geostatistical models offer a robust methodology for identifying determinants of the disease distribution and for predicting infection risk and burden at high spatial scales (Diggle et al., 1998). These models have been widely used in assessing

the relationship between helminth infection with demographic, environmental, and socioeconomic predictors, at sub-national (Raso et al., 2005, 2006a), national (Clements et al., 2009a), or regional scales (Clements et al., 2010; Schur et al., 2011a, 2013). In the Americas, high resolution, geostatistical, model-based risk estimates have been obtained for the whole continent (Chammartin et al., 2013b) as well as for Brazil (Scholte et al., 2013). A key issue in geostatistical modelling is the selection of the predictors. Most of the variable selection methods in geostatistical applications rely on standard methods, such as stepwise regression or bivariate associations that are appropriate for non-spatial data (Brooker et al., 2002; Raso et al., 2005). However, ignoring spatial correlation leads to incorrect estimates of the statistical significance of the predictors included in the model. Recently, Bayesian variable selection has been introduced in geostatistical disease mapping (Giardina et al., 2012; Schur et al., 2013).

The purpose of this paper was to map the geographical distribution of *A. lumbricoides*, *T. trichiura*, and hookworm in Bolivia, and to estimate the risk, number of infected school-aged children, and the costs related to treatment interventions in the country. Survey data were extracted from published and unpublished sources. Bayesian geostatistical models were employed using rigorous variable selection procedures.

5.2 Methods

5.2.1 Disease data

Data on the prevalence of soil-transmitted helminth infection were extracted from the global neglected tropical diseases (GNTD) database (www.gntd.org) (Hürlimann et al., 2011; Schur et al., 2011a, 2013; Stensgaard et al., 2013; Chammartin et al., 2013b). The GNTD database is an open-access platform consisting of georeferenced survey data pertaining to schistosomiasis, soil-transmitted helminthiasis, and other neglected tropical diseases. Surveys are identified through systematic searches of electronic databases such as PubMed and ISI Web of Knowledge with no restriction of publication date or language. Our search strategy, including data quality appraisal, is summarised in Table 5.1.

5.2.2 Environmental, socioeconomic, and population data

A total of 40 environmental and socioeconomic variables were considered in our analysis. Environmental variables included 19 interpolated climatic data from weather stations related to temperature and precipitation, vegetation proxies such as the enhanced vegetation index (EVI) and normalised difference vegetation index (NDVI), altitude, land cover,

Table 5.1: Search strategy identification of soil-transmitted helminth infection prevalence survey data in Bolivia.

Keywords	Exclusion criteria	Quality control measures
Bolivi* AND helminth* (OR ascari*, OR trichur*, OR hookworm, OR necator, OR ankylostom*,OR ancylostom*,OR strongy*, OR hymenolepis*,OR toxocara*, OR enterobius*, OR geohelminth*, OR nematode)	Hospital-based study; case-control study (except control group); clinical trials (except baseline); drug-efficacy study (except placebo group); displaced population (travellers, military, expatriates, nomads); population treated for the infection during the past year; unclear location of the survey; sample size < 10	Double check of each entry; search and elimination of duplicates; recalculation of prevalence; verification in Google Maps that point level coordinates correspond to human settlement

as well as information on soil acidity and soil moisture. Various unsatisfactory basic needs (UBN) poverty indicators related to adequate housing material, insufficient housing space, inadequate services of water and sewer systems and inadequate health attention were used as proxies of poverty. In addition, human development index (HDI) and infant mortality rate (IMR) were considered as alternative poverty measures. Impact of direct human influence on ecosystems was accounted by human influence index (HII). Population density and the proportion of school-aged children (age: 5–14 years), were used to estimate treatment needs and costs of intervention. Sources of the variables, together with their spatial and temporal resolution, are summarised in Table 5.2.

For prediction purposes, a 5×5 km spatial resolution grid was created. Environmental data available at 1×1 km spatial resolution, were averaged over their closest neighbours. Soil acidity, soil moisture, and infant mortality rate were linked to the prediction pixel with the closest distance. UBN and HDI were re-scaled by assigning to each grid pixel the value of the administrative unit they belong to. Re-scaling was performed in ArcMap version 10.0 (Environmental Systems Research Institute; Redlands, CA, USA).

Table 5.2: Data sources and properties of the predictors explored to model soil-transmitted helminth infection risk in Bolivia.

Data type	Source	Date	Temporal resolution	Spatial resolution
19 climatic variables related to temperature and precipitation	WorldClim[1]	1950–2000	-	1 km
Altitude	SRTM[2]	2000	-	1 km
Land cover	MODIS/Terra[3]	2000–2011	Yearly	1 km
EVI/NDVI	MODIS/Terra[3]	2000–2011	16 days	1 km
Soil acidity/soil moisture	ISRIC-WISE[4]	1960–2000	-	10 km
Unsatisfactory basic need (UBN)	Census[5]	2001	10 years	Municipality
Infant mortality rate (IMR)	CIESIN[6]	2005	Yearly	5 km
Human influence index (HII)	LTW[7]	2005	-	1 km
Human development index (HDI)	PAHO[8]	2005	-	Municipality
Population density	GPWFE[4]	2010	-	10 km
School-aged children proportion	IDB[10]	2010	-	Country

[1] WorldClim Global Climate database version1.4. Available at: http://www.worldclim.org/ (accessed: March 2012).

[2] Shuttle Radar Topography Mission (SRTM). Available at: http://www.worldclim.org/ (accessed: March 2012).

[3] Moderate Resolution Imaging Spectroradiometer (MODIS). Available at: https://lpdaac.usgs.gov/ (accessed: December 2012).

[4] ISRIC-WISE database (WISE3). Available at:http://www.isric.org/ (accessed: December 2012).

[5] Instituto nacional de estadística, 2001 census. Available at: http://www.ine.gob.bo/(http://www.ine.gob.bo/ (accessed: March 2012).

[6] 2005 Global subnational infant mortality rates, Center for International Earth Science Information Network (CIESIN). CIESIN, Palisades, NY, USA. Available at: http://www.ciesin.columbia.edu/povmap/ds_global.html (accessed: March 2012).

[7] Last of the Wild Data version 2, 2005 (LTW-2): Global Human Footprint Dataset (Geographic), Wildlife Conservation (WCS) and Center for International Earth Science Information Network (CIESIN). Available at: http://www.ciesin.org/wildareas/ (accessed: March 2012).

[8] Pan American Health Organization; personal communication.

[9] Gridded Population of the World: future estimates (GPWFE): Center for International Earth Science Information Network (CIESIN), UN Food and Agriculture Programme (FAO) and Centro Internacional de Agricultura Tropical (CIAT). Available at: http://sedac.ciesin.columbia.edu/gpw (accessed: December 2012).

[10] International Data Base (IDB) United States census Bureau. Available at http://www.census.gov/population/international/ (accessed: March 2012).

Geostatistical model

Disease survey data are typically binomially distributed and modelled *via* a logistic regression. More precisely, let Y_i, n_i, and p_i be the number of infected individuals, the number of individuals screened, and the prevalence or risk of infection at location i, respectively, such as $Y_i \sim Bn(n_i, p_i)$. Spatial correlation is taken into account by introducing location-specific parameters φ_i that are considered as unobserved latent data from a stationary spatial Gaussian process. We modelled a temporal trend, the selected predictors (i. e. environmental and socioeconomic factors) X_i and φ_i on the logit scale: $logit(p_i) = X_i^T \underline{\beta} + \varphi_i$. The temporal trend was modelled by a binary variable T_i indicating whether a survey was carried out before or from 1995 onwards. We assumed that $\underline{\varphi} \sim MVN(\underline{0}, \Sigma)$ with variance-covariance matrix Σ. Geographical correlation was modelled by an isotropic exponential correlation function of distance, i. e. $\Sigma_{cd} = \sigma_{sp}^2 exp(-\rho d_{cd})$, where d_{cd} is the Euclidean distance between locations c and d, σ_{sp}^2 is the geographical variability known as the partial sill, and ρ is a smoothing parameter controlling the rate of correlation decay. The geographic dependency (range) was defined as the minimum distance at which spatial correlation between locations is less than 5% and is calculated by $3/\rho$. To facilitate model fit, the model was formulated using a Bayesian framework of inference. Vague normal prior distributions $\underline{\beta} \sim N(\underline{0}, \sigma^2 I)$ were adopted for the regression coefficients, an inverse-gamma distribution $\sigma_{sp}^2 \sim IG(a_{\sigma_{sp}^2}, b_{\sigma_{sp}^2})$ was chosen for the variance σ_{sp}^2, and a gamma distribution was assumed for the spatial decay ρ, $\rho \sim G(a_\rho, b_\rho)$.

Geostatistical variable selection

Bayesian stochastic search variable selection (George and McCulloch, 1993) was performed to select the most important predictors among the 40 socioeconomic and environmental predictors, while taking into account the spatial correlation in the data. Predictors were either standardised or categorised if they presented a non-linear bivariate association with the observed helminthiasis prevalence (on the logit scale). Furthermore, we considered a spike and slab prior distribution for the regression coefficients (Scheipl et al., 2012), which improves convergence properties of the Markov chain Monte Carlo (MCMC) simulation and allows selection of blocks of covariates such as categorical ones. In addition, we assessed correlation between the predictors and forced the model to choose only one (or none) predictor among those highly correlated (i. e. absolute value of Pearson's correlation coefficient larger than 0.9). The geostatistical variable selection explores all possible models and the final model is the one presenting the highest posterior probability.

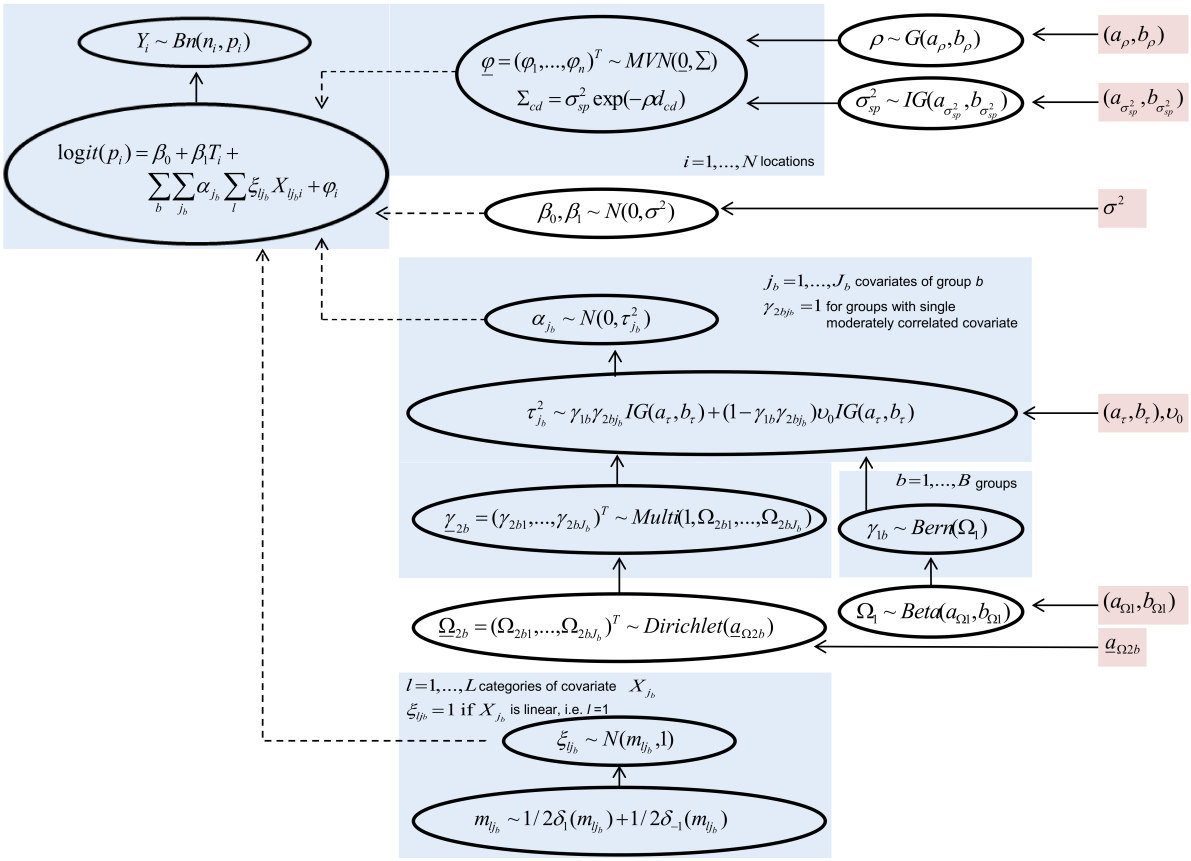


Figure 5.1: Acyclic graph of the geostatistical variable selection. Stochastic and logical nodes are represented as ellipses. Dashed arrows are logical links and straight line arrows are stochastic dependencies. Fixed parameters of the prior distributions are highlighted in pink.

The geostatistical variable selection specification is summarised in Figure 5.1. In particular, predictors were classified into 19 groups b, $(b = 1, ..., 19)$, depending on their mutual correlations. Thirteen predictors that were only moderately correlated with any other predictors were separated into single variable groups. Highly correlated predictors were divided into six groups, each containing 3–8 variables X_{j_b}, $j_b = 1, ..., J_b$. The regression coefficients are defined as the product of an overall contribution α_{j_b} of the predictor X_{j_b} and the effect ξ_{lj_b} of each of its elements (i. e. categories), X_{lj_b}, $l = 1, ..., L$ categories (excluding baseline) of the predictor X_{j_b}. We assigned a spike and slab prior (Scheipl et al., 2012; Chammartin et al., 2013a), which is a scaled normal mixture of inverse-gamma to α_{j_b}, that is $\alpha_{j_b} \sim N(0, \tau^2_{j_b})$, where $\tau^2_{j_b} \sim \gamma_{1b}\gamma_{2bj_b} IG(a_\tau, b_\tau) + (1 - \gamma_{1b}\gamma_{2bj_b}) \upsilon_0 IG(a_\tau, b_\tau)$. a_τ and b_τ are

fixed parameters of non-informative inverse-gamma distribution, while υ_0 is a small constant shrinking α_{j_b} to zero when the predictor is excluded. The presence or absence of the predictors is defined by the product of two indicators γ_{1b} and $\underline{\gamma}_{2b} = (\gamma_{2b1}, ..., \gamma_{2bJ_b})^T$, where γ_{1b} determines the presence or absence of the group b in the model and γ_{2bj_b}, $j_b = 1, ..., J_b$, allows selection of a single predictor within the group. A Bernoulli and a multinomial prior distribution are assigned to γ_{1b} and $\underline{\gamma}_{2b}$, respectively, such as $\gamma_{1b} \sim Bern(\Omega_1)$ and $\underline{\gamma}_{2b} \sim Multi(1, \Omega_{2b1}, ..., \Omega_{2bJ_b})$ with inclusion probabilities Ω_1 and $\underline{\Omega}_{2b}$. To allow greater flexibility in estimating model size, these probabilities are considered as hyper-parameters having non-informative beta and Dirichlet distributions. A mixture of two Gaussian distributions is assumed for ξ_{lj_b}, $\xi_{lj_b} \sim N(m_{lj_b}, 1)$, $m_{lj_b} \sim 1/2\delta_1(m_{lj_b}) + 1/2\delta_{-1}(m_{lj_b})$, which shrinks ξ_{lj_b} towards $|1|$ (multiplicative identity). For predictors moderately correlated, γ_{2bj_b} is fixed to 1, while the effect of linear predictors is only defined by an overall contribution of α.

To complete model specification, the spatial random effect $\underline{\varphi}$ is modelled as defined in the previous subsection and a vague normal distribution is assigned to the constant term of the model. The subset of variables included in the models with the highest posterior probabilities identified the final models.

Implementation details

We considered the following values for the parameters of the prior distributions: $\sigma^2 = 100$, $(a_\rho, b_\rho) = (0.01, 0.01)$, $(a_{\sigma^2_{sp}}, b_{\sigma^2_{sp}}) = (2.01, 1.01)$, $(a_\tau, b_\tau) = (5, 25)$, $(a_{\Omega 1}, b_{\Omega 1}) = (1, 1)$, $\underline{a}_{\Omega 2b} = (1, ..., 1)$, and $\upsilon_0 = 0.00025$.

MCMC simulations were used to estimate model parameters. For variable selection, a burn-in of 50,000 iterations was performed and another 50,000 iterations were run to identify the model with the highest posterior probability. For each infection, the best geostatistical model was fitted with one chain sampler and a burn-in of 5,000 iterations. Convergence was assessed after an average of 50,000 iterations using the Raftery and Lewis (1992) diagnostics. A posterior sample of 1,000 values was used for validation purposes and for prediction at un-sampled locations. Prediction was carried out using Bayesian kriging (Diggle et al., 1998) over a grid of 26,519 pixels of 5×5 km spatial resolution. The median and standard deviation of the predicted posterior distribution were plotted to produce smooth risk maps together with their uncertainty. Analyses were implemented in WinBUGS 14 (Imperial College and Medical Research Council; London, UK), while R

version 2.7.2 (R Core Team, 2014) was used for predictions. Non-spatial explorative statistical analyses were performed in Stata version 10.0 (Stata Corporation; College Station, USA).

Model validation

Models were fitted on a random training sample of 39 locations for *A. lumbricoides* and *T. trichiura*, and 37 locations for hookworm. Model validation was performed on the remaining 10 test locations (around 20% of the total locations). The predictive performance was calculated by the proportion of test locations being correctly predicted within the kth Bayesian credible interval (BCI) of the posterior predictive distribution (limited by the lower and upper quantiles $BCI^l_{i(k)}$ and $BCI^u_{i(k)}$, respectively), where k indicates the probability coverage of the interval as: $\frac{1}{10}\sum_{i=1}^{10}\min(I(BCI^l_{i(k)} < p_i), I(BCI^u_{i(k)} > p_i))$. The higher the number of test locations within the narrowest and smallest coverage BCI, the better the model predictive ability.

Treatment needs and estimated costs

The number of infected school-aged children was calculated for each pixel from the geostatistical model-based estimated risk and the population density. According to guidelines put forward by the World Health Organization (WHO), all school-aged children should be treated twice a year in high-risk communities (prevalence of any soil-transmitted helminth infection $\geq$ 50%) and once every year in low-risk communities (prevalence of any soil-transmitted helminth infection between 20% and 50%). Large-scale preventive chemotherapy is not recommended in areas where prevalence is less than 20%; indeed treatment should be delivered on a case-by-case basis in such areas (WHO, 2006). We estimated the number of albendazole or mebendazole treatments needed during one year in the school-aged population, considering different units at which levels of risk were determined (i. e. pixel, municipality, province, and department). Hence, we followed the same methodology as for estimating annualised praziquantel needs against schistosomiasis (Schur et al., 2012). To calculate the cost of a school-based deworming programme in Bolivia, the estimated number of treatments was multiplied by an average unit cost equivalent to US$ 0.25, which includes additional expenses for training, drug distribution, and administration (Bitran et al., 2009; IADB, PAHO and SABIN, 2011).

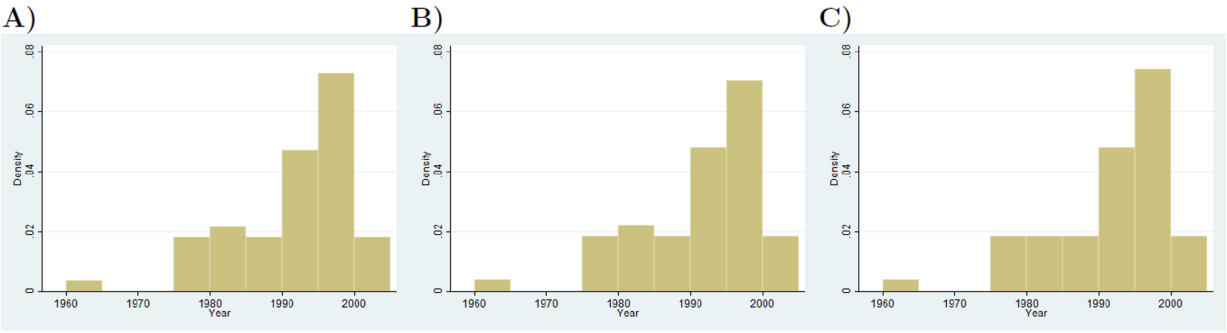


Figure 5.2: Frequency distribution of the survey periods in Bolivia for *A. lumbricoides* (A), *T. trichiura* (B), and hookworm (C).

5.3 Results

Seven out of 59 identified peer-reviewed publications reported soil-transmitted helminth infection prevalence data in Bolivia (Basset et al., 1986; Cancrini et al., 1989; Esteban et al., 1997; Cancrini et al., 1998; Flores et al., 2001; Tanner, 2005; Benefice et al., 2006). For the current investigation, additional data were obtained from a 2006 report of the Ministry of Health (MoH) in Bolivia (Mollinedo and Prieto, 2006).

We obtained relevant prevalence data for *A. lumbricoides*, *T. trichiura*, and hookworm for 49, 49, and 47 survey locations, respectively, covering the period from 1960 to 2010. The frequency distribution of the surveys, stratified by helminth species, is given in Figure 5.2. Six surveys out of 49 were reported at municipality level (administrative level 3) and were assigned to the centroid of their municipality. The remaining 43 locations were reported at school or village level and were therefore considered as point data. Most of the studies (71%) explicitly screened school-aged children (the remaining studies are either referring to entire populations or provide no information on the age range of the participants). With regard to the diagnosis of soil-transmitted helminthiasis, 47% of the studies used the WHO-recommended Kato-Katz technique (Montresor et al., 1998), whereas in 21 locations the diagnostic approach was not stated, and in five locations other diagnostic techniques were utilised.

Table 5.3 summarises, for each helminth species, the three best models resulting from the geostatistical variable selection. For *A. lumbricoides*, the model based on precipitation of the wettest quarter has the highest posterior probability of 42.2%. For *T. trichiura* the best model included altitude (posterior probability = 10.1%), while for hookworm, the model with the highest posterior probability (10.2%) included the minimum temperature

Table 5.3: Variables selected by the geostatistical variable selection approach.

	***A. lumbricoides* infection**			***T. trichiura* infection**			**Hookworm infection**		
Group 1									
Home with indoor plumbing[1]	0	0	0	0	0	0	0	0	0
People with drinking water home[1]	0	0	0	0	0	0	0	0	0
Pipe network	0	0	0	0	0	0	0	0	0
Population with high quality of life	0	0	0	0	0	0	0	0	0
Population with UBN	0	0	0	0	0	0	0	0	0
Population with sanitation home	0	0	0	0	X	0	0	0	0
Group 2									
Population with material UBN	0	0	0	0	0	0	0	0	0
Population with low quality of life	0	0	0	0	0	0	0	0	0
Group 3									
Min temperature coldest month[1,2]	0	0	0	0	0	0	X	0	X
Altitude	0	0	0	X	0	0	0	0	0
Annual temperature	0	0	0	0	X	0	0	0	0
Max temperature warmest month	0	0	0	0	0	0	0	0	0
Temperature wettest quarter	0	0	0	0	0	0	0	X	0
Temperature driest quarter	0	0	0	0	0	0	0	0	0
Temperature warmest quarter	0	0	0	0	0	0	0	0	0
Temperature coldest quarter	0	0	0	0	0	0	0	0	0
Group 4									
Temperature annual range[3]	0	0	0	0	0	0	0	0	0
Temperature diurnal range	0	0	0	0	0	X	0	0	0
Group 5									
Annual precipitation[1,2,3]	0	0	0	0	0	0	0	0	0
Precipitation wettest month[1,2]	0	0	0	0	0	0	0	0	0
Precipitation wettest quarter[1,2]	X	0	0	0	0	0	0	0	0
Precipitation driest month[2,3]	0	0	0	0	0	0	0	0	0
Precipitation driest quarter[2]	0	0	0	0	0	0	0	0	0
Precipitation warmest quarter[3]	0	0	0	0	0	0	0	0	0
Precipitation coldest quarter[2]	0	0	0	0	0	0	0	0	0
Group 6									
Enhanced vegetation index	0	0	0	0	0	0	0	0	0
Normalised difference vegetation index	0	0	0	0	0	0	0	0	X

Continued on next page

	***A. lumbricoides* infection**			***T. trichiura* infection**			**Hookworm infection**		
Moderately isolated									
Soil acidity[1,3]	0	0	X	0	0	0	0	0	0
Precipitation seasonality[1,3]	0	0	0	0	0	0	0	0	0
Soil moisture[2]	0	0	0	0	0	0	0	0	0
Isothermality	0	0	0	0	0	0	0	0	0
Temperature seasonality	0	0	0	0	0	0	0	0	0
Human influence index	0	0	0	0	0	0	0	0	0
Infant mortality rate	0	0	0	0	0	0	0	0	0
Human development index	0	0	0	0	0	0	0	0	0
Population with education UBN	0	0	0	0	0	0	0	0	0
Population with overcrowding UBN	0	0	0	0	0	0	0	0	0
Population with sanitation UBN	0	0	0	0	0	0	0	0	0
Population with light home	0	0	0	0	X	0	0	0	0
Unemployment rate	0	0	0	0	0	0	0	0	0
Posterior probability (%)	42.2	5.9	2.9	10.1	6.0	5.2	10.2	4.7	2.0

X: selected; 0: not selected; [1] categorised for *A. lumbricoides*; [2] categorised for *T. trichiura*; [3] categorised for hookworm.
The best three models selected by the geostatistical variable selections are presented for each soil-transmitted helminth species, together with their posterior probabilities.

during the coldest month. Results of the geostatistical logistic regressions, together with estimates of the bivariate non-spatial associations, are presented in Table 5.4. Precipitation of the wettest quarter above 400 mm had a positive effect on the odds of *A. lumbricoides* infection risk; hookworm infection risk was positively associated to the minimum temperature during the coldest month, and the higher the altitude, the lower the odds of *T. trichiura* infection. Although the risk of infection with the three helminth species decreased after 1995, this effect was not important in the spatial models as reflected by the 95% BCI of the odds ratio estimates. Figures 5.3, 5.4, and 5.5 show the geographical distribution of the predicted risks for each of the three soil-transmitted helminth species before and after 1995, the corresponding standard deviation of the predictive distribution and the raw survey data. Maps of all predictors involved in the final geostatistical models are shown in Figure 5.6. Bolivia presents generally a lower risk of soil-transmitted helminthiasis in the south-western part of the country, where high altitude brings unsuitable climatic conditions for the development of the parasites. For the three soil-transmitted helminth infections, the maps of the posterior standard deviation reflect the pattern of the predicted risk. However, we note that for hookworm, where the spatial correlation is more

important (spatial range estimated to 128.4 km), the standard deviation was also low in areas surrounding the survey locations, suggesting less uncertainty in the estimation of the spatial random effect in the neighbourhood of observed data. Figure 5.7 shows that the risks of *A. lumbricoides*, *T. trichiura* and hookworm infection are correctly predicted within 95% BCIs for 90%, 90%, and 80%, respectively.

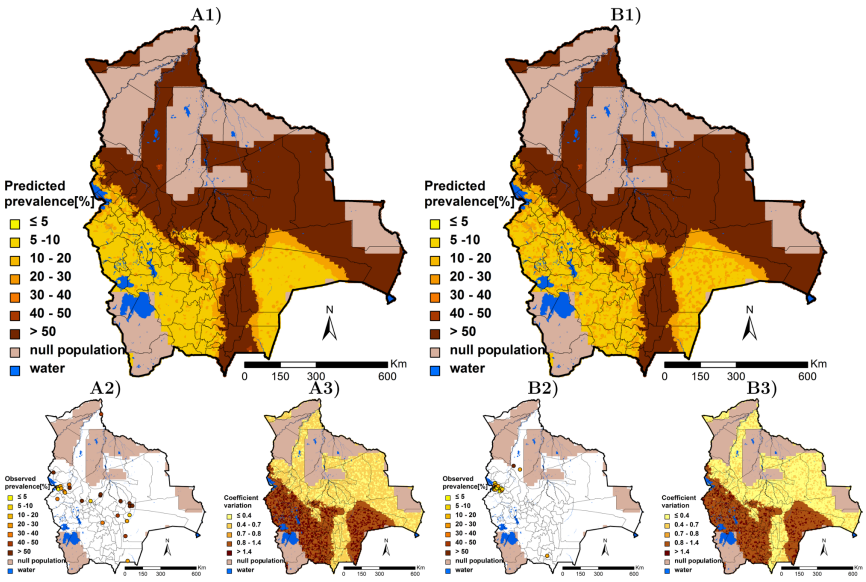


Figure 5.3: *Ascaris lumbricoides* infection risk in Bolivia. The maps show the situation before 1995 (A) and from 1995 onwards (B), and provide estimates of the geographical distribution of the infection (1), the observed prevalence (2), and the coefficient of variation (3).

Table 5.4: Parameter estimates of non-spatial bivariate and Bayesian geostatistical logistic models with environmental and socio-economic predictors.

	Bivariate non-spatial		**Geostatistical model**	
***A. lumbricoides* infection**	**OR**	**95% CI**	**OR**	**95% BCI**
Survey period				
< 1995	1.00		1.00	
≥ 1995	0.26	(0.24; 0.29)*	0.94	(0.64; 1.42)
Precipitation wettest quarter (mm)				
< 350	1.00		1.00	
350–400	1.42	(1.23, 1.66)*	1.32	(0.56; 2.81)
≥ 400	12.25	(10.95, 13.70)	12.52	(5.05; 25.56)*
			Median	**95% BCI**
σ^2			1.11	(0.72; 2.00)
Range (km)			9.2	(1.3; 63.0)
***T. trichiura* infection**	**OR**	**95%5 CI**	**OR**	**95% BCI**
Survey period				
< 1995	1.00		1.00	
≥ 1995	0.33	(0.29; 0.37)*	0.85	(0.55; 1.30)
Altitude	0.33	(0.31, 0.36)*	0.37	(0.26; 0.56)*
			Median	**95% BCI**
σ^2			1.29	(0.77; 2.23)
Range (km)			28.7	(3.2; 80.2)
Hookworm infection	**OR**	**95%5 CI**	**OR**	**95% BCI**
Survey period				
< 1995	1.00		1.00	
≥ 1995	0.45	(0.41; 0.50)*	0.72	(0.12; 4.19)
Minimum temperature coldest month	6.25	(5.81, 6.72) *	11.35	(5.00; 22.20)*
			Median	**95% BCI**
σ^2			3.07	(1.50; 7.44)
Range (km)			128.4	(39.8; 387.5)

* Significant based on 95% CI or 95% BCI.
OR: odds ratio; 95% CI: lower and upper bound of a 95% confidence interval; 95% BCI: lower and upper bound of a 95% Bayesian credible interval.

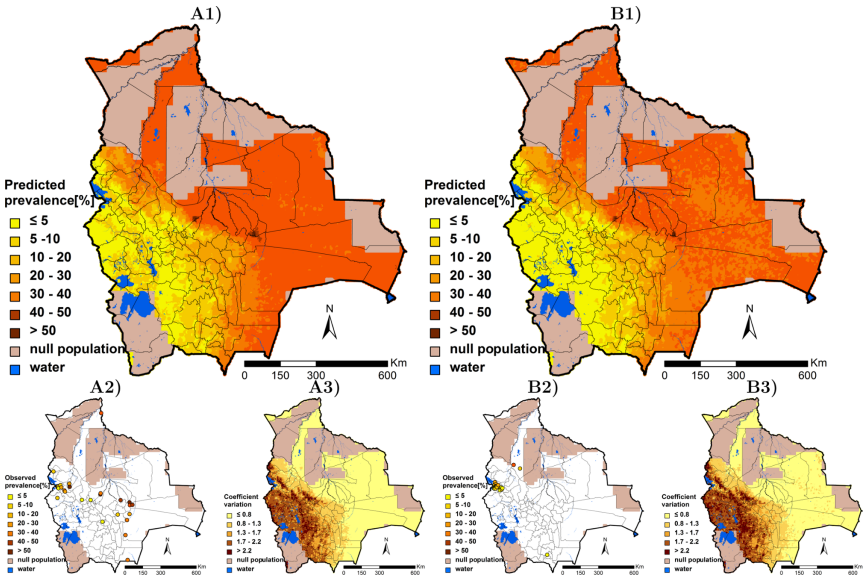


Figure 5.4: *Trichuris trichiura* infection risk in Bolivia. The maps show the situation before 1995 (A) and from 1995 onwards (B), and provide estimates of the geographical distribution of the infection (1), the observed prevalence (2), and the coefficient of variation (3).

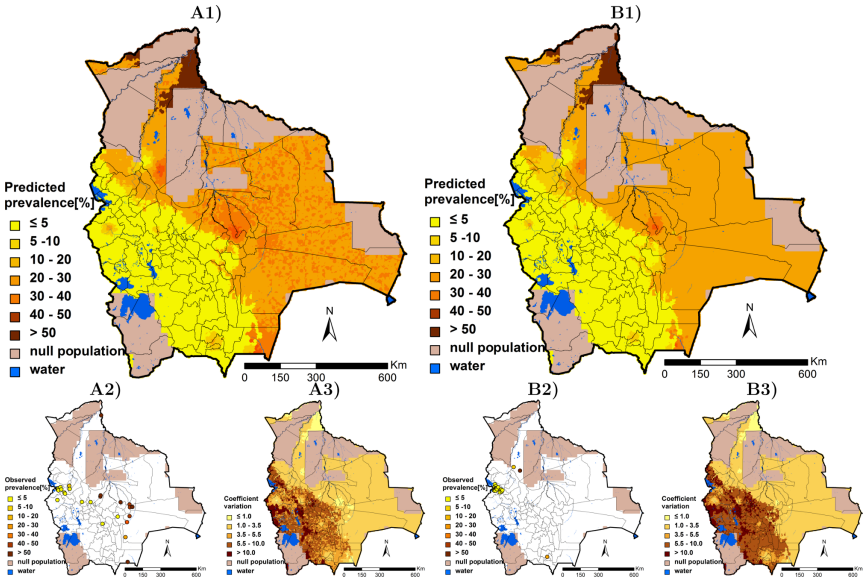


Figure 5.5: Hookworm infection risk in Bolivia. The maps show the situation before 1995 (A) and from 1995 onwards (B), and provide estimates of the geographical distribution of the infection (1), the observed prevalence (2), and the coefficient of variation (3).

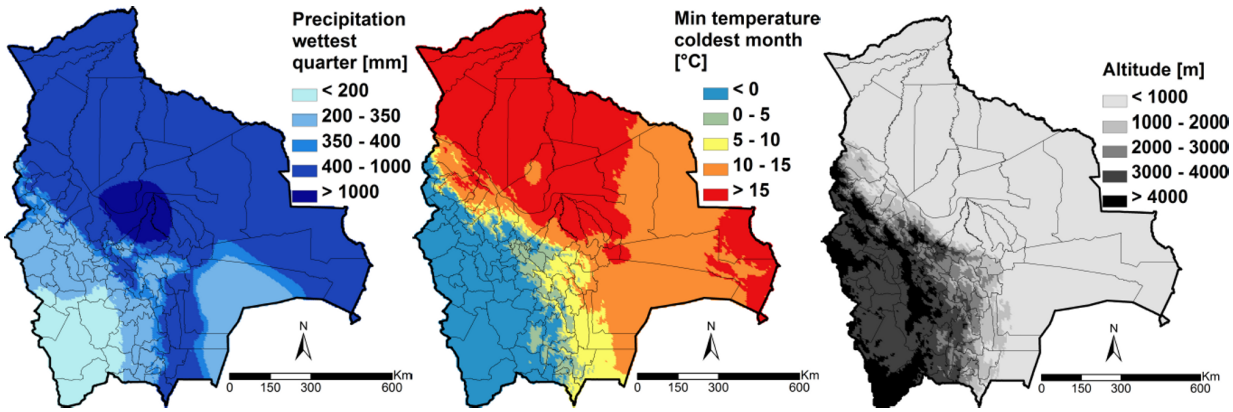


Figure 5.6: Major climatic zones and spatial distribution of the remotely sensed predictors in Bolivia.

Table 5.5 shows the total amount of treatment required on a yearly basis and the associated cost when the calculation is based on soil-transmitted helminth infection risk estimates, aggregated to various administrative levels. The estimated number of children targeted increases from 1,481,605 to 2,180,101, depending on the administrative level at which the risk is aggregated. However, the number of treatments required remains quite stable, indicating large spatial heterogeneity of the infection risk within the units. Model-based predictions and estimates of number of school-aged children infected with the three soil-transmitted helminth species, aggregated at province and country level, are presented in the Appendix 5.6 (Table 5.6 and 5.7). The estimated prevalence for *A. lumbricoides*, *T. trichiura*, and hookworm infection is 38.0%, 19.3%, and 11.4%, respectively. Taking the three soil-transmitted helminth species together, we estimate that 48.4% of the school-aged population is infected with at least one species, assuming independence of the three soil-transmitted helminth infections. The highest number of school-aged children needing treatment is concentrated in the densely populated Andrés Ibáñez province, while the highest risk for the three soil-transmitted helminths taken together is predicted for the Vaca Díez province.

5.4 Discussion

We present spatially explicit estimates of the risk and number of school-aged children infected with the three common soil-transmitted helminths in Bolivia using a rigorous geostatistical variable selection approach. Survey data were extracted from the literature,

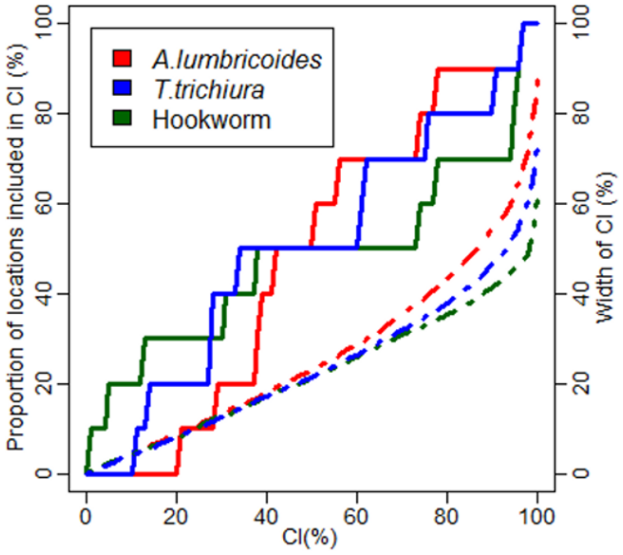


Figure 5.7: Proportion of locations with observed prevalence falling within credible intervals of the posterior predictive distribution with probability coverage varying from 1% to 100%.

Table 5.5: Yearly estimation of school-aged children needing preventive chemotherapy against soil-transmitted helminthiases in Bolivia.

	5×5 **km**	**Municipality**	**Province**	**Department**
Number of children targeted	1,481,605	1,749,136	1,907,658	2,180,101
Number of treatment delivered	2,894,936	2,868,016	2,847,604	3,013,413
Cost (US$)	723,734	717,003	711,901	753,353

Estimates are based on prevalence predicted at pixels of 5×5 km resolution and aggregated over different administrative levels.

georeferenced, and made public *via* the open-access GNTD database. Our study also identified important data needs and gaps. For example, most of the surveys were conducted along the sub-Andean region. On the other hand, only few survey locations were available in the less densely populated highlands and in the northern tropical areas. Rigorous geostatistical variable selection methods have been used to identify environmental and socioeconomic determinants that govern the distribution of soil-transmitted helminth infection in Bolivia. The country, nestled between the high Andean peaks (on the West) and the Amazon forest (on the East), presents specific ecological characteristics that shape helminth cycles in a complex way. High altitude and diverse topography, as well as the paucity of weather stations in remote areas can introduce interpolation bias in the climatic

factors used in our analysis (Hijmans et al., 2005). Bayesian variable selection helped in identifying the potential factors influencing the geographical distribution of the three common soil-transmitted helminth species. Our methodology enabled us to explore all possible models arising from 40 climatic and socioeconomic predictors, while accounting for spatial correlation in the data.

The parameterisation of the prior distribution of the regression coefficients as developed in this manuscript selects the best predictors among highly correlated ones, while addressing non-linearity. The selected predictors are plausible in terms of helminth biology, ecology, and epidemiology. Indeed, the distribution of *A. lumbricoides* was positively associated with precipitation above 400 mm during the wettest month. High humidity is related with faster development of parasite eggs in the free environment. Low humidity, on the other hand, can cease embryonation of *A. lumbricoides* (Otto, 1929; Spindler, 1929). The positive association between the minimum temperature of the coldest month and the prevalence of hookworm reflects inhibition of the development of the eggs by hostile cold temperatures (Smith et al., 1990; Pullan and Brooker, 2012). The preventive effect of high altitude on *T. trichiura* infection risk has already been highlighted and explained by subsequent unfavourable temperature, which limits the transmission (Bundy and Cooper, 1989). The three soil-transmitted helminth infection risks did not decrease significantly over time and we are unsure whether Bolivia has implemented integrated control measures. In the absence of preventive chemotherapy and/or sanitation improvement, environmental contamination is considerable, which may explain our observations of fairly constant infection rates over time (Hotez et al., 2009; Smits, 2009).

The transmission of soil-transmitted helminthiasis occurs via contaminated food or fingers (*A. lumbricoides* and *T. trichiura*), or through the skin by walking on larvae-infested soil (hookworm). People living in poor conditions are more exposed due to their living conditions, the lack of access to clean water, sanitation, and health facilities (De Silva et al., 2003). Hence, we would have expected soil-transmitted helminth infections to be associated with some of the socioeconomic factors investigated, such as the ones related to sanitation (Ziegelbauer et al., 2012). However, none of the socioeconomic variables were picked up by our geostatistical variable selection approach. This may indicate that our socioeconomic proxies were not able to capture the socioeconomic disparities across the country when aggregated at district or municipality scales. Historical data are aggregated over villages or larger areas and they are rarely available at household level. Often variation in socioeconomic status is larger within rather than between locations, and hence, it may

be harder for socioeconomic data to explain geographical differences.

Bolivian soil also exhibits specific characteristics such as presence of salt and soil compactation arising from livestock farming, which may affect the transmission of soil-transmitted helminths. In our analysis, we explored different soil predictors, including land cover, the vegetation indices EVI and NDVI, soil acidity and soil moisture. However, these factors failed to explain the distribution of the infection risks.

The population of Bolivia is mainly concentrated in and around the three main cities La Paz, Santa Cruz, and Cochabamba, where large parts of the country are uninhabited. The absence of human hosts breaks parasite life cycles. Thus, although environmental conditions may be suitable for parasite survival, there is no risk of transmission. To avoid potential misinterpretation, we clearly delineate areas where no humans live.

The predicted risk maps for the three common soil-transmitted helminth species in Bolivia should be interpreted with caution, particularly for areas characterised by only sparse survey data or poor coverage. Sample design is not optimised regarding the surveyed population; 29% of the data did not report the survey type (school-aged, community-based) and might bias the raw prevalence, as it is widely acknowledged that school-aged children are at higher risk of soil-transmitted helminths, particularly *A. lumbricoides* and *T. trichiura*, than their older counterparts (Albonico et al., 1993). Slightly less than half of the surveys stated the use of the WHO-recommended Kato-Katz technique for soil-transmitted helminth diagnosis (Montresor et al., 1998; Speich et al., 2010). Heterogeneity in the data regarding the sensitivities and specificities of the diagnostic methods might introduce measurement errors in the raw prevalence data. Furthermore, a zero hookworm prevalence was reported for 60% of the survey data. While these data suggest the non-endemicity of hookworm, the diagnostic approach might have underestimated the "true" prevalence due to diagnostic dilemmas (Knopp et al., 2008; Bergquist et al., 2009). Indeed, single Kato-Katz thick smears, low intensity infections, and delays in stool processing compromise sensitivity, particularly for hookworm diagnosis (Dacombe et al., 2007; Krauth et al., 2012). Giardina et al. (2012) developed a zero-inflated binomial geostatistical model to estimate malaria burden when data contain a high proportion of zeros. This model could be adopted for soil-transmitted helminth infection and implemented in Bolivia as soon as more survey data become available. In addition, data in the literature usually report on hookworm prevalence, without differentiation of the species (*A. duodenale* and *N. americanus*). It would be interesting to analyse the two species separately, as they may have different ecological preferences.

Our study indicates that in Bolivia almost half (48.4%) of the population is infected with at least one of the three common soil-transmitted helminths. Our empirical-based estimates suggested that a total of 2,868,016 annualised treatments are required for preventive chemotherapy targeting school-aged children at the level of the municipalities. This estimate is higher than the one previously reported in the country (4,774,672 treatments for a 5-year campaign (Bitran et al., 2009; IADB, PAHO and SABIN, 2011)). Population dynamic models (Anderson and May, 1985; Chan et al., 1994; Anderson et al., 2012) could be used to predict the effect of preventive chemotherapy on the epidemiological pattern of the three common soil-transmitted helminths, to evaluate the community effectiveness of the programme and to plan the duration of control interventions.

5.5 Conclusion

In the framework of a preventive chemotherapy strategy, reliable maps of the distribution of infection risk and disease burden are needed to enhance cost-effectiveness of the interventions. Our high resolution estimates are based on existing data and their scarcity may raise doubts on the value of modelling of the disease distribution. However, soil-transmitted helminth infections are driven by environmental factors and, in the absence of interventions, the existing data can establish the relation between the risk of infection and climate. Hence, the risk maps produced are able to identify areas of high infection. Validation indicated that the models had good predictive ability. We therefore believe that the estimated maps can provide important inputs in the sampling design of a national survey by indicating the areas requiring more surveys. Hence, a coherent and optimally designed national survey is warranted to more accurately estimate the distribution and the number of people at risk of infection, so that preventive chemotherapy and other control measures can be optimally targeted.

Acknowledgements

The authors are grateful for financial support of the Pan American Health Organization (PAHO) and the UBS Optimus Foundation. RGCS received further financial support from the Swiss Brazilian Joint Research Programme (BSJRP 011008). We thank the reviewers for providing valuable comments in an earlier version of the manuscript.

5.6 Appendix

Table 5.6: Population-adjusted prevalence with the three common soil-transmitted helminth infections, stratified by province and by country, for the period 1995 onwards, based on 2010 population estimates with 95% Bayesian credible interval (BCI).

	Model-based prevalence (%)			
	Ascaris lumbricoides infection	*Trichuris trichiura* infection	Hookworm infection	Soil-transmitted helminth infection*
Province				
Abel Iturralde	55.4 (45.6; 64.4)	34.5 (22.8; 48.1)	23.2 (7.3; 41.6)	69.9 (61.1; 78.6)
Alonso de Ibáñez	13.3 (8.8; 22.6)	6.8 (3.7; 12.4)	3.0 (0.5; 11.4)	20.9 (14.3; 29.4)
Andrés Ibáñez	55.8 (45.7; 63.8)	34.2 (24.2; 46.3)	20.7 (5.2; 46.6)	70.4 (61.6; 80.0)
Aniceto Arce	41.3 (34.8; 47.6)	22.9 (15.3; 31.4)	6.9 (1.4; 20.5)	52.5 (44.9; 59.5)
Antonio Quijarro	13.2 (8.9; 22.6)	6.0 (3.2; 11.3)	3.5 (0.7; 12.0)	20.3 (14.9; 28.8)
Arani	33.6 (26.7; 42.0)	16.5 (11.5; 23.7)	5.6 (1.3; 15.6)	42.2 (34.9; 49.5)
Aroma	13.1 (8.7; 22.9)	6.7 (3.9; 11.6)	4.0 (0.8; 15.3)	21.3 (14.7; 30.9)
Arque	20.7 (14.8; 28.2)	6.7 (3.6; 12.4)	2.4 (0.5; 9.5)	26.8 (19.8; 34.4)
Atahuallpa	13.1 (8.9; 23.1)	6.8 (3.6; 12.3)	3.7 (0.7; 13.1)	20.9 (15.4; 30.3)
Ayopaya	55.4 (44.8; 65.2)	12.2 (8.0; 18.3)	7.6 (2.1; 18.7)	60.3 (50.6; 69.5)
Bautista Saavedra	42.7 (34.9; 55.9)	11.5 (7.5; 17.4)	7.7 (1.8; 18.1)	48.8 (40.3; 60.8)
Belisario Boeto	37.9 (30.7; 45.6)	18.2 (11.8; 26.5)	7.2 (1.6; 19.4)	48.1 (39.2; 56.6)
Bernardino Bilbao	36.9 (28.3; 45.9)	9.6 (4.9; 16.9)	2.4 (0.3; 9.8)	42.0 (32.9; 51.6)
Burnet O'Connor	38.2 (32.0; 44.6)	21.2 (14.5; 29.7)	7.6 (1.5; 20.3)	50.2 (42.4; 57.5)
Capinota	36.9 (27.8; 47.7)	10.2 (4.4; 19.3)	1.2 (0.1; 7.0)	42.8 (31.9; 54.3)
Carangas	13.0 (8.8; 23.0)	7.0 (3.8; 12.3)	4.8 (1.0; 16.7)	22.0 (15.6; 32.3)
Carrasco	49.6 (41.4; 57.6)	26.4 (18.1; 36.4)	17.0 (5.4; 34.9)	61.4 (53.5; 69.7)
Cercado (Cochabamba)	18.9 (10.4; 29.6)	3.5 (0.8; 15.5)	0.1 (0.0; 1.2)	22.0 (12.0; 37.0)
Cercado (Oruro)	13.0 (8.8; 22.7)	6.6 (3.6; 11.3)	2.9 (0.5; 11.9)	20.5 (15.2; 29.2)
Cercado (Tarija)	14.8 (10.0; 23.4)	12.3 (7.0; 20.4)	10.2 (3.4; 18.5)	29.8 (21.1; 40.1)
Cercado (El Beni)	55.4 (45.7; 64.6)	35.7 (23.3; 50.5)	24.6 (7.5; 43.6)	70.7 (61.0; 79.2)
Chapare	48.5 (37.9; 59.5)	10.8 (6.9; 17.9)	4.5 (1.3; 10.0)	53.2 (42.3; 62.9)
Charcas	45.3 (37.8; 52.9)	11.3 (6.8; 17.5)	3.4 (0.7; 10.9)	50.2 (42.8; 58.0)
Chayanta	30.1 (24.9; 35.9)	8.3 (4.9; 14.2)	3.8 (0.8; 13.0)	36.3 (30.3; 43.6)
Chiquitos	46.0 (38.2; 54.2)	35 (22.7; 47.9)	24.4 (7.2; 47.0)	66.0 (55.1; 76.3)
Cordillera	29.5 (24.4; 34.6)	32.2 (22.1; 45.7)	22.9 (6.9; 43.7)	56.6 (46.2; 67.5)
Cornelio Saavedra	13.7 (9.4; 22.2)	9.2 (5.3; 15.2)	3.8 (0.8; 13.8)	23.3 (17.1; 32.0)
Daniel Campos	13.1 (8.8; 22.8)	7.0 (3.9; 12.3)	3.7 (0.8; 13.4)	21.3 (15.4; 30.1)
Eduardo Avaroa	13.1 (8.8; 23.3)	5.8 (3.2; 11.3)	3.5 (0.8; 12.9)	20.3 (14.2; 28.4)
Eliodoro Camacho	22.2 (16.1; 35.6)	6.1 (3.2; 12.0)	2.9 (0.6; 9.8)	28.0 (20.5; 40.5)
Esteban Arce	23.9 (17.4; 33.5)	9.6 (4.7; 16.6)	1.2 (0.1; 5.1)	30.6 (23.2; 40.2)
Eustaquio Méndez	14.1 (9.5; 23.9)	11.5 (6.9; 18.6)	8.9 (2.0; 18.5)	28.0 (19.7; 39.0)
Federico Román	55.1 (45.6; 64.0)	35.0 (22.7; 48.5)	38.2 (15.8; 56.9)	74.2 (64.4; 82.5)
Florida	20.4 (14.0; 30.9)	22.6 (14.1; 33.5)	16.7 (4.0; 39.3)	43.6 (30.9; 58.3)
Franz Tamayo	51.0 (42.1; 58.4)	21.2 (14.9; 29.7)	14.8 (4.9; 30.5)	60.3 (52.2; 68.1)
Germán Jordán	19.1 (8.8; 33.3)	9.6 (2.5; 22.4)	0.6 (0.0; 5.8)	26.6 (12.2; 43.9)
Gran Chaco	33.1 (28.0; 38.5)	31.2 (20.7; 43.5)	18.6 (5.2; 38.9)	57.5 (47.9; 68.6)
Gualberto Villarroel	13.1 (8.3; 23.2)	7.1 (3.6; 12.9)	5.7 (1.0; 21.3)	22.4 (14.3; 34.4)
Hernando Siles	55.2 (46.4; 64.5)	22.7 (15.9; 32.6)	6.0 (1.6; 16.4)	62.9 (54.7; 71.3)
Ichilo	54.4 (44.4; 64.8)	31 (20.1; 43.9)	24.1 (7.7; 48.7)	69.3 (58.5; 79.5)
Ignacio Warnes	56.5 (45.0; 67.6)	39.1 (24.5; 54.5)	19.5 (5.1; 45.5)	72.7 (60.5; 83.0)
Ingavi	14.9 (10.7; 22.7)	7.0 (4.2; 11.7)	1.9 (0.4; 6.5)	21.4 (16.4; 29.3)
Inquisivi	36.3 (30.3; 41.7)	9.6 (6.0; 14.8)	4.3 (1.0; 14.1)	41.5 (35.4; 48.6)
Jaime Zudáñez	14.6 (10.5; 23.0)	14.1 (9.2; 20.7)	4.3 (1.0; 13.5)	27.9 (21.3; 35.9)
José Ballivián	55.2 (45.9; 64.4)	34.3 (22.6; 48.0)	23.7 (10.4; 37.5)	70.3 (61.9; 77.8)
J.M. Avilés	14.5 (9.1; 25.8)	12.8 (7.2; 23.1)	8.8 (2.3; 20.0)	29.0 (19.4; 42.8)
J.M. Linares	13.2 (8.7; 22.9)	8.7 (5.2; 14.7)	4.5 (0.9; 14.3)	22.9 (16.2; 31.3)
J.M. de Velasco	55.2 (46.2; 64.2)	34.3 (22.7; 47.6)	24.0 (7.4; 43.4)	70.2 (61.3; 78.0)

Continued on next page

	Model-based prevalence (%)			
	Ascaris lumbricoides infection	*Trichuris trichiura* infection	Hookworm infection	Soil-transmitted helminth infection*
J.A. de Padilla	31.9 (26.1; 39.0)	14.3 (8.9; 21.2)	4.5 (1.0; 14.5)	41.5 (34.5; 49.4)
Ladislao Cabrera	13.1 (8.9; 22.7)	7.2 (4.0; 12.8)	4.7 (1.1; 17.7)	22.0 (16.2; 32.1)
Larecaja	50.1 (41; 59.6)	20.2 (13.7; 28.9)	11.2 (2.9; 26.3)	58.8 (50.0; 67.1)
Litoral	13.2 (8.2; 23.2)	7.2 (4.0; 12.8)	5.0 (0.9; 17.8)	22.1 (14.7; 33.9)
Loayza	19.6 (14.9; 27.0)	9.0 (5.1; 15.4)	2.7 (0.5; 10.4)	27.2 (21.1; 35.2)
Los Andes	12.5 (8.0; 21.5)	5.5 (3.5; 9.8)	0.3 (0.1; 1.5)	17.3 (12.2; 26.3)
Luis Calvo	38.1 (32.1; 44.2)	28.0 (18.8; 39.3)	14.7 (4.2; 30.4)	57.6 (48.5; 67.2)
Madre de Dios	55.3 (45.9; 65.1)	36.0 (23.2; 49.4)	24.1 (7.1; 43.2)	70.6 (61.2; 79.1)
Mamoré	55.4 (45.4; 64.6)	36.0 (23.6; 49.7)	27.0 (8.9; 46.8)	71.4 (62.0; 79.9)
Manco Kapac	55.9 (36.7; 70.9)	6.2 (2.6; 14.9)	1.2 (0.1; 9.8)	58.0 (38.5; 73.3)
M.M. Caballero	35.1 (29.5; 41.2)	19.3 (13.0; 27.1)	9.6 (2.6; 21.8)	47.3 (40.2; 54.5)
Manuripi	55.4 (45.6; 64.5)	35.1 (23.1; 49.1)	25.9 (8.3; 45.1)	71.2 (61.7; 79.7)
Marbán	55.3 (45.5; 64.3)	35.1 (23; 49.4)	24.6 (7.4; 44.1)	70.3 (61.4; 79.1)
Mizque	35.7 (28.6; 41.8)	12.0 (7.6; 18.4)	2.9 (0.7; 10.3)	42.5 (35.0; 49.6)
Modesto Omiste	13.1 (8.3; 22.7)	8.1 (4.4; 13.7)	4.5 (0.8; 14.2)	22.5 (15.6; 31.5)
Moxos	55.5 (46.1; 64.9)	34.9 (22.5; 48.9)	23.7 (7.1; 43.2)	70.1 (60.3; 79.0)
Muñecas	42.3 (34.2; 51.7)	10.5 (6.3; 16.8)	4.2 (0.8; 12.4)	47.2 (38.8; 56.8)
Narciso Campero	16.0 (11.8; 24.4)	16.4 (11.2; 23.7)	4.3 (1.0; 13.1)	30.6 (24.5; 38.7)
Nor Chichas	13.0 (8.7; 22.4)	9.0 (5.2; 14.2)	5.3 (1.1; 17.0)	23.6 (17.3; 32.5)
Nor Cinti	14.9 (11.0; 23.3)	10.9 (6.8; 16.2)	5.3 (1.1; 15.9)	26.1 (20.0; 33.8)
Nor Lípez	13.1 (9.0; 22.7)	6.7 (3.8; 12.1)	3.5 (0.7; 12.3)	20.7 (15.3; 29.4)
Nor Yungas	53.0 (42.1; 64.5)	19.9 (13.4; 28.4)	9.5 (2.7; 23.5)	61.8 (51.2; 72.5)
Ñuflo de Chávez	54.8 (45.4; 63.6)	34.2 (22.4; 48.6)	24.8 (7.7; 45.0)	70.2 (60.7; 78.6)
Obispo Santistevan	55.2 (46.2; 64.7)	35.2 (23.0; 49.2)	24.3 (7.5; 46.7)	70.6 (61.1; 80.2)
Omasuyos	15.2 (10.5; 22.5)	8.1 (5.0; 12.5)	0.5 (0.1; 2.7)	22.0 (16.1; 28.5)
Oropeza	33.5 (26.8; 40.4)	11.9 (7.3; 18.4)	3.1 (0.6; 12.3)	41.4 (34.2; 48.7)
Pacajes	13.4 (9.1; 22.3)	6.1 (3.4; 10.7)	5.0 (1.0; 16.9)	22.0 (15.7; 31.5)
Pantaleón Dalence	13.0 (7.5; 24.0)	5.8 (2.8; 12.3)	0.6 (0.1; 5.1)	18.3 (11.4; 29.2)
P.D. Murillo	23.3 (18.8; 29.5)	9.1 (6.1; 13.1)	2.8 (0.6; 7.8)	29.0 (24.1; 35.2)
Poopó	13.0 (8.6; 23.1)	6.4 (3.3; 12.5)	3.1 (0.5; 12.8)	20.3 (14.0; 29.79
Punata	13.5 (6.2; 26.0)	9.9 (4.0; 21.4)	0.9 (0.1; 7.3)	22.1 (11.2; 40.2)
Quillacollo	46.8 (36.2; 56.5)	7.0 (2.7; 13.9)	0.9 (0.1; 4.7)	50.1 (39.2; 60.3)
Rafael Bustillo	13.1 (8.4; 22.8)	6.5 (3.5; 11.4)	3.6 (0.7; 12.0)	20.6 (14.2; 29.9)
Sajama	13.1 (8.9; 23.2)	6.2 (3.3; 11.1)	3.5 (0.6; 14.6)	20.5 (14.8; 30.1)
Sara	56.4 (44.9; 67.0)	38.6 (26.5; 53.6)	20.3 (5.7; 47.4)	72.8 (61.7; 83.4)
Saucarí	13.1 (8.3; 24.5)	7.4 (3.8; 13.4)	4.9 (1.0; 17.9)	22.4 (15.0; 34.19
Sud Chichas	12.9 (8.9; 22.5)	7.7 (4.3; 13.1)	5.0 (1.0; 14.7)	22.4 (16.0; 31.2)
Sud Cinti	21.8 (17.6; 28.0)	15.1 (10.3; 21.6)	6.7 (1.3; 17.1)	34.1 (27.4; 41.59
Sud Lípez	13.0 (8.9; 22.5)	4.7 (2.4; 9.6)	0.7 (0.2; 4.8)	17.5 (12.9; 26.4)
Sud Yungas	36.9 (30.5; 44.4)	15.8 (10.8; 23.0)	6.3 (2.2; 13.1)	44.7 (37.8; 52.5)
Tapacarí	40.7 (31.8; 49.4)	7.2 (3.4; 13.1)	2.4 (0.4; 10.8)	44.7 (35.4; 53.9)
Tomás Frías	13.1 (8.7; 23.0)	6.3 (3.3; 12.1)	3.8 (0.9; 14.0)	20.8 (14.3; 30.3)
Tomina	36.9 (30.7; 43.0)	16.7 (11.0; 24.3)	6.3 (1.6; 16.8)	46.6 (39.9; 53.8)
Vaca Díez	55.2 (46.5; 63.9)	35.8 (23.7; 49.5)	45.1 (19.2; 64.3)	76.0 (65.5; 84.9)
Vallegrande	40.9 (34.0; 47.9)	19.4 (13.0; 27.4)	7.0 (1.5; 19.5)	51.0 (44.0; 58.1)
Yacuma	55.3 (46.0; 63.9)	35.8 (23.2; 49.7)	33.9 (13.1; 51.1)	73.0 (63.9; 81.2)
Yamparáez	16.1 (9.4; 27.8)	10.9 (6.4; 17.9)	3.3 (0.5; 13.0)	26.3 (17.7; 37.4)
Country				
Bolivia	38 (32.4; 43.4)	19.3 (14.2; 25.9)	11.4 (3.6; 23.7)	48.4 (43.1; 54.4)

* Calculated under the assumption of independenc of *A. lumbricoides*, *T.trichiura* and hookworm infections.

Table 5.7: Estimated number of infected children (5–14 years old) with the three common soil-transmitted helminth infections, stratified by province and by country, for the period 1995 onwards, based on 2010 population estimates with 95% Bayesian credible interval (BCI).

	Number of school-aged children infected*			
	Ascaris lumbricoides infection	*Trichuris trichiura* infection	Hookworm infection	Soil-transmitted helminth infection**
Province				
Abel Iturralde	2,627 (2,164; 3,057)	1,635 (1,083; 2,283)	1,099 (347; 1,972)	3,316 (2,899; 3,728)
Alonso de Ibáñez	872 (574; 1,477)	443 (241; 811)	199 (35; 743)	1,363 (934; 1,925)
Andrés Ibáñez	180,358 (147,634; 206,402)	110,743 (78,296; 149,638)	66,809 (16,757; 150,726)	227,630 (199,126; 258,702)
Aniceto Arce	5,691 (4,796; 6,564)	3,149 (2,111; 4,321)	958 (193; 2,822)	7,229 (6,187; 8,194)
Antonio Quijarro	1,742 (1,180; 2,982)	798 (424; 1,487)	469 (98; 1,579)	2,686 (1,968; 3,806)
Arani	4,031 (3,201; 5,036)	1,984 (1,374; 2,841)	676 (155; 1,876)	5,058 (4,192; 5,941)
Aroma	3,119 (2,061; 5,458)	1,606 (918; 2,773)	960 (194; 3,651)	5,066 (3,499; 7,366)
Arque	3,389 (2,425; 4,615)	1,093 (587; 2,028)	398 (79; 1,552)	4,380 (3,233; 5,625)
Atahuallpa	487 (332; 859)	252 (135; 458)	137 (25; 487)	777 (574; 1,127)
Ayopaya	12,922 (10,443; 15,214)	2,835 (1,875; 4,271)	1,779 (498; 4,350)	14,061 (11,798; 16,212)
Bautista Saavedra	2,309 (1,889; 3,027)	624 (407; 942)	418 (97; 979)	2,639 (2,179; 3,288)
Belisario Boeto	1,011 (819; 1,218)	487 (315; 707)	193 (44; 518)	1,286 (1,047; 1,513)
Bernardino Bilbao	1,195 (918; 1,485)	309 (159; 546)	78 (11; 318)	1,360 (1,065; 1,671)
Burnet O'Connor	5,269 (4,412; 6,149)	2,923 (1,998; 4,090)	1,050 (211; 2,798)	6,914 (5,840; 7,928)
Capinota	5,373 (4,043; 6,950)	1,483 (647; 2,816)	170 (20; 1,020)	6,228 (4,646; 7,901)
Carangas	1,221 (825; 2,152)	656 (357; 1,157)	454 (91; 1,569)	2,064 (1,466; 3,024)
Carrasco	18,629 (15,537; 21,646)	9,907 (6,790; 13,657)	6,367 (2,043; 13,112)	23,066 (20,089; 26,191)
Cercado (Cochabamba)	9,445 (5,170; 14,784)	1,740 (408; 7,739)	58 (4; 579)	10,983 (5,991; 18,456)
Cercado (Oruro)	6,022 (4,068; 10,472)	3,065 (1,655; 5,230)	1,327 (249; 5,483)	9,453 (7,012; 13,483)
Cercado (Tarija)	4,217 (2,845; 6,681)	3,518 (1,998; 5,816)	2,909 (958; 5,270)	8,517 (6,019; 11,460)
Cercado (El Beni)	10,815 (8,912; 12,594)	6,958 (4,554; 9,858)	4,808 (1,460; 8,506)	13,790 (11,898; 15,444)
Chapare	73,219 (57,168; 89,884)	16,348 (10,484; 27,060)	6,756 (2,021; 15,029)	80,295 (63,844; 94,967)
Charcas	6,758 (5,638; 7,885)	1,688 (1,013; 2,610)	511 (109; 1,619)	7,488 (6,380; 8,653)
Chayanta	10,064 (8,328; 12,005)	2,786 (1,629; 4,751)	1,276 (284; 4,355)	12,136 (10,137; 14,572)
Chiquitos	21,253 (17,674; 25,037)	16,188 (10,489; 22,152)	11,288 (3,319; 21,720)	30,509 (25,482; 35,240)
Cordillera	10,494 (8,694; 12,327)	11,454 (7,886; 16,279)	8,168 (2,443; 15,574)	20,168 (16,455; 24,016)
Cornelio Saavedra	2,741 (1,891; 4,447)	1,840 (1,057; 3,039)	760 (155; 2,769)	4,670 (3,415; 6,403)
Daniel Campos	281 (189; 489)	150 (84; 264)	80 (16; 287)	456 (331; 645)
Eduardo Avaroa	2,090 (1,407; 3,715)	932 (516; 1,808)	563 (127; 2,058)	3,237 (2,256; 4,518)
Eliodoro Camacho	2,519 (1,831; 4,045)	695 (366; 1,364)	332 (66; 1,112)	3,173 (2,327; 4,592)
Esteban Arce	3,288 (2,393; 4,596)	1,324 (649; 2,284)	164 (18; 706)	4,200 (3,181; 5,524)
Eustaquio Méndez	2,233 (1,510; 3,792)	1,818 (1,091; 2,952)	1,417 (324; 2,943)	4,439 (3,128; 6,191)
Federico Román	6,896 (5,712; 8,012)	4,377 (2,840; 6,066)	4,778 (1,972; 7,121)	9,285 (8,055; 10,324)
Florida	4,238 (2,914; 6,404)	4,678 (2,931; 6,959)	3,458 (837; 8,150)	9,054 (6,405; 12,085)
Franz Tamayo	3,313 (2,738; 3,800)	1,378 (969; 1,930)	965 (321; 1,984)	3,922 (3,392; 4,429)
Germán Jordán	677 (312; 1,180)	341 (89; 794)	23 (1; 204)	943 (433; 1,557)
Gran Chaco	12,549 (10,624; 14,595)	11,850 (7,857; 16,507)	7,053 (1,974; 14,749)	21,822 (18,161; 26,007)
Gualberto Villarroel	947 (602; 1,686)	514 (260; 937)	414 (75; 1,546)	1,628 (1,035; 2,492)
Hernando Siles	4,443 (3,734; 5,190)	1,826 (1,280; 2,625)	482 (132; 1,317)	5,067 (4,401; 5,745)
Ichilo	28,867 (23,539; 34,362)	16,467 (10,639; 23,306)	12,788 (4,097; 25,829)	36,733 (31,045; 42,139)
Ignacio Warnes	22,228 (17,704; 26,608)	15,387 (9,642; 21,439)	7,683 (2,004; 17,912)	28,582 (23,817; 32,647)
Ingavi	4,608 (3,333; 7,051)	2,174 (1,303; 3,623)	593 (120; 2,004)	6,628 (5,086; 9,070)
Inquisivi	6,712 (5,600; 7,710)	1,778 (1,110; 2,732)	801 (180; 2,600)	7,678 (6,549; 8,984)
Jaime Zudáñez	1,306 (936; 2,058)	1,261 (819; 1,852)	386 (88; 1,206)	2,494 (1,903; 3,208)
José Ballivián	10,950 (9,111; 12,783)	6,809 (4,491; 9,513)	4,708 (2,058; 7,439)	13,944 (12,270; 15,437)
J.M. Avilés	2,048 (1,285; 3,635)	1,811 (1,022; 3,263)	1,247 (325; 2,823)	4,093 (2,743; 6,041)
J.M. Linares	1,483 (974; 2,570)	983 (583; 1,653)	501 (106; 1,605)	2,575 (1,823; 3,513)
J.M. de Velasco	11,240 (9,406; 13,069)	6,981 (4,627; 9,705)	4,895 (1,500; 8,833)	14,300 (12,485; 15,887)
J.A. de Padilla	2,071 (1,691; 2,533)	928 (580; 1,377)	293 (66; 939)	2,693 (2,239; 3,203)

Continued on next page

	Number of school-aged children infected*			
	Ascaris lumbricoides infection	*Trichuris trichiura* infection	Hookworm infection	Soil-transmitted helminth infection**
Ladislao Cabrera	530 (360; 915)	291 (160; 516)	190 (42; 714)	889 (652; 1,295)
Larecaja	66,801 (54,729; 79,558)	26,941 (18,273; 38,599)	15,004 (3,861; 35,094)	78,370 (66,738; 89,516)
Litoral	124 (77; 219)	68 (38; 121)	47 (9; 168)	208 (139; 320)
Loayza	4,427 (3,364; 6,103)	2,039 (1,163; 3,473)	614 (120; 2,347)	6,157 (4,774; 7,973)
Los Andes	3,994 (2,570; 6,864)	1,759 (1,121; 3,137)	111 (22; 470)	5,539 (3,906; 8,427)
Luis Calvo	2,377 (2,002; 2,758)	1,744 (1,171; 2,453)	918 (263; 1,896)	3,590 (3,024; 4,195)
Madre de Dios	1,902 (1,577; 2,240)	1,238 (799; 1,700)	829 (246; 1,486)	2,428 (2,104; 2,718)
Mamoré	2,478 (2,031; 2,888)	1,610 (1,053; 2,224)	1,207 (396; 2,093)	3,192 (2,770; 3,573)
Manco Kapac	3,945 (2,587; 5,004)	440 (181; 1,049)	84 (7; 694)	4,092 (2,718; 5,172)
M.M. Caballero	2,831 (2,381; 3,324)	1,557 (1,048; 2,188)	774 (209; 1,762)	3,821 (3,244; 4,403)
Manuripi	1,607 (1,323; 1,869)	1,018 (671; 1,423)	751 (241; 1,308)	2,066 (1,789; 2,310)
Marbán	8,160 (6,717; 9,478)	5,182 (3,385; 7,291)	3,629 (1,091; 6,509)	10,368 (9,052; 11,666)
Mizque	3,709 (2,978; 4,345)	1,244 (787; 1,914)	299 (69; 1,069)	4,419 (3,637; 5,160)
Modesto Omiste	960 (608; 1,660)	592 (323; 1,000)	330 (60; 1,037)	1,645 (1,140; 2,302)
Moxos	7,380 (6,136; 8,632)	4,642 (2,990; 6,510)	3,158 (946; 5,743)	9,331 (8,020; 10,508)
Muñecas	2,770 (2,240; 3,383)	689 (413; 1,101)	276 (52; 815)	3,091 (2,543; 3,717)
Narciso Campero	1,985 (1,470; 3,030)	2,042 (1,397; 2,952)	533 (125; 1,627)	3,798 (3,043; 4,811)
Nor Chichas	961 (645; 1,653)	665 (385; 1,044)	389 (85; 1,257)	1,741 (1,276; 2,398)
Nor Cinti	2,252 (1,668; 3,532)	1,648 (1,033; 2,452)	811 (164; 2,411)	3,958 (3,033; 5,124)
Nor Lípez	718 (496; 1,248)	366 (206; 665)	190 (41; 674)	1,139 (839; 1,617)
Nor Yungas	21,583 (17,155; 26,242)	8,095 (5,435; 11,574)	3,885 (1,101; 9,575)	25,148 (20,862; 29,507)
Ñuflo de Chávez	25,700 (21,275; 29,827)	16,019 (10,519; 22,788)	11,632 (3,601; 21,122)	32,915 (28,483; 36,843)
Obispo Santistevan	18,466 (15,468; 21,634)	11,783 (7,692; 16,467)	8,119 (2,521; 15,621)	23,614 (20,440; 26,832)
Omasuyos	3,774 (2,599; 5,562)	2,007 (1,228; 3,099)	135 (13; 657)	5,451 (3,992; 7,070)
Oropeza	17,185 (13,745; 20,716)	6,110 (3,755; 9,413)	1,574 (315; 6,331)	21,239 (17,542; 25,002)
Pacajes	2,276 (1,540; 3,796)	1,033 (576; 1,821)	855 (174; 2,880)	3,733 (2,670; 5,350)
Pantaleón Dalence	780 (451; 1,441)	351 (169; 740)	36 (5; 304)	1,097 (684; 1,753)
P.D. Murillo	54,432 (44,011; 68,903)	21,185 (14,271; 30,607)	6,454 (1,495; 18,242)	67,673 (56,221; 82,183)
Poopó	759 (503; 1,347)	376 (193; 727)	180 (32; 747)	1,184 (816; 1,733)
Punata	761 (350; 1,465)	558 (223; 1,204)	53 (4; 411)	1,245 (632; 2,264)
Quillacollo	26,104 (20,213; 31,482)	3,916 (1,517; 7,749)	495 (67; 2,641)	27,916 (21,835; 33,602)
Rafael Bustillo	1,881 (1,208; 3,290)	935 (499; 1,640)	516 (107; 1,731)	2,974 (2,050; 4,314)
Sajama	472 (319; 833)	222 (120; 400)	127 (21; 526)	736 (532; 1,084)
Sara	33,743 (26,889; 40,095)	23,061 (15,858; 32,034)	12,172 (3,392; 28,381)	43,552 (36,912; 49,864)
Saucarí	786 (501; 1,470)	442 (230; 804)	296 (59; 1,077)	1,344 (903; 2,049)
Sud Chichas	1,302 (896; 2,266)	772 (431; 1,319)	508 (97; 1,478)	2,253 (1,616; 3,144)
Sud Cinti	1,363 (1,105; 1,753)	948 (645; 1,352)	419 (84; 1,072)	2,135 (1,719; 2,601)
Sud Lípez	288 (197; 499)	105 (53; 214)	16 (3; 108)	389 (288; 587)
Sud Yungas	16,409 (13,577; 19,740)	7,021 (4,794; 10,247)	2,784 (961; 5,805)	19,878 (16,818; 23,327)
Tapacarí	8,080 (6,311; 9,807)	1,421 (667; 2,602)	479 (85; 2,137)	8,873 (7,028; 10,703)
Tomás Frías	5,484 (3,668; 9,637)	2,665 (1,381; 5,059)	1,587 (364; 5,874)	8,727 (5,998; 12,716)
Tomina	3,452 (2,871; 4,019)	1,562 (1,029; 2,271)	588 (149; 1,571)	4,358 (3,730; 5,034)
Vaca Díez	22,559 (18,993; 26,125)	14,651 (9,691; 20,253)	18,415 (7,848; 26,297)	31,067 (26,787; 34,705)
Vallegrande	2,674 (2,218; 3,126)	1,264 (846; 1,789)	455 (101; 1,273)	3,334 (2,871; 3,797)
Yacuma	4,597 (3,826; 5,313)	2,978 (1,927; 4,129)	2,818 (1,091; 4,248)	6,067 (5,311; 6,748)
Yamparáez	1,841 (1,072; 3,180)	1,243 (730; 2,051)	379 (59; 1,486)	3,008 (2,026; 4,274)
Country				
Bolivia	941,453 (801,990; 1,073,112)	478,856 (352,584; 642,099)	281,448 (89,989; 586,020)	1,198,182 (1,067,216; 1,347,507)

* Calculated based on the 5–14 years old population for 2010.

** Calculated under the assumption of independenc of *A. lumbricoides*, *T.trichiura* and hookworm infections.

Chapter 6

Spatio-temporal distribution of soil-transmitted helminth infections in Brazil

Chammartin F.[1,2], Guimarães L.H.[3], Scholte R.G.C.[4], Bavia M.E.[5], Utzinger J.[1,2] and Vounatsou P.[1,2]

[1] Swiss Tropical and Public Health Institute, Basel, Switzerland
[2] University of Basel, Basel, Switzerland
[3] Immunology service, Hospital Edgard Santos, Federal University of Bahia, Salvador, Bahia, Brazil
[4] Coordenação Geral de Hanseníase e Doenças em Eliminação, Secretaria de Vigilância em Saúde, Brasília, Distrito Federal, Brazil
[5] Preventive Medicine Department, Federal University of Bahia, Salvador, Bahia, Brazil

This paper has been published in *Parasites & Vectors* 2014, 7: 440.

Abstract

Background: In Brazil, preventive chemotherapy targeting soil-transmitted helminthiasis is being scaled-up. Hence, spatially explicit estimates of infection risks providing information about the current situation are needed to guide interventions. Available high-resolution national model-based estimates either rely on analyses of data restricted to a given period of time, or on historical data collected over a longer period. While efforts have been made to take into account the spatial structure of the data in the modelling approach, little emphasis has been placed on the temporal dimension.

Methods: We extracted georeferenced survey data on the prevalence of infection with soil-transmitted helminths (i. e. *Ascaris lumbricoides*, hookworm and *Trichuris trichiura*) in Brazil from the Global Neglected Tropical Diseases (GNTD) database. Selection of the most important predictors of infection risk was carried out using a Bayesian geostatistical approach and temporal models that address non-linearity and correlation of the explanatory variables. The spatial process was estimated through a predictive process approximation. Spatio-temporal models were built on the selected predictors with integrated nested Laplace approximation using stochastic partial differential equations.

Results: Our models revealed that, over the past 20 years, the risk of soil-transmitted helminth infection has decreased in Brazil, mainly because of the reduction of *A. lumbricoides* and hookworm infections. From 2010 onwards, we estimate that the infection prevalences with *A. lumbricoides*, hookworm and *T. trichiura* are 3.6%, 1.7% and 1.4%, respectively. We also provide a map highlighting municipalities in need of preventive chemotherapy, based on a predicted soil-transmitted helminth infection risk in excess of 20%. The need for treatments in the school-aged population at the municipality level was estimated at 1.8 million doses of anthelminthic tablets per year.

Conclusion: The analysis of the spatio-temporal aspect of the risk of infection with soil-transmitted helminths contributes to a better understanding of the evolution of risk over time. Risk estimates provide the soil-transmitted helminthiasis control programme in Brazil with useful benchmark information for prioritising and improving spatial and temporal targeting of interventions.

6.1 Introduction

The nematode worms *Ascaris lumbricoides*, *Trichuris trichiura* and the two hookworm species *Ancylostoma duodenale* and *Necator americanus* are commonly referred to as soil-transmitted helminths (Bethony et al., 2006). These nematodes parasitise the human intestine and might lead to chronic infections with clinical consequences that undermine health of affected populations (Bethony et al., 2006; Hall et al., 2008; Hotez et al., 2008a). The World Health Organization (WHO) advocates a global control strategy against major helminthiases, emphasising preventive chemotherapy targeting high-risk communities, in combination with health education and sanitation improvement whenever resources allow (WHO, 2012).

Soil-transmitted helminthiases are of considerable public health concern in tropical and subtropical countries, where climatic conditions and poverty-related behaviours favour their transmission (Pullan and Brooker, 2012; Pullan et al., 2014). South America is not spared (Chammartin et al., 2013b; Scholte et al., 2013). In Brazil, deworming campaigns were carried out covering up to 60% of the population but interventions have been interrupted in 2005 (WHO, 2011b, 2012), partially because of the decentralization of the programme (Gabrielli et al., 2013). Currently, WHO estimates that 9 million school-aged children in Brazil require preventive chemotherapy and anthelminthic administration of albendazole has been re-implemented in 2013 (WHO, 2012).

Spatial targeting of the population requiring preventive chemotherapy and other interventions is essential to implement tailored and cost-effective control measures. Bayesian geostatistical models are used to establish a statistical relationship between observed prevalence and environmental and socioeconomic risk factors, and predict the risk at unobserved locations, while accounting for spatial heterogeneity through spatially structured random effects (Diggle et al., 1998). These models have been widely applied to model soil-transmitted helminth risk at different scales (Raso et al., 2006a; Clements et al., 2010; Chammartin et al., 2013b; Lai et al., 2013). They are highly parameterised, and therefore estimation of model parameters relies on Markov chain Monte Carlo (MCMC) sampling methods. However, inference requires multiple inversions of the spatially structured variance-covariance matrix and MCMC methods are known to be computationally intensive. Thus, for large datasets, spatial process estimation can rely on low-rank approximation, such as the predictive process (Banerjee et al., 2008; Chammartin et al., 2013b).

By incorporating a temporal trend into the model, changes of the disease risk and

pattern over time can be studied (Chammartin et al., 2013b,c; Lai et al., 2013). A temporal trend assumes that the infection risk changes over time by a certain amount, which is constant across space. However, the underlying latent spatial process might also vary over time. Bayesian formulations introduced by Knorr-Held (2000) allow accounting for space-time interaction with an effect that is spatio-temporally structured through its precision matrix. Hence, such spatio-temporal models are able to estimate the spatial variation with time. The spatio-temporal aspect of helminthiases risk is an under-explored issue, mainly because of computational challenges in estimating highly parameterised models with MCMC algorithms. However, recent developments in Bayesian inference with integrated nested Laplace approximation (INLA) (Rue et al., 2009) using stochastic partial differential equations (SPDEs) (Lindgren et al., 2011) offer new opportunities for accurate fit of complex models at reasonable computational cost and time (Karagiannis-Voules et al., 2013).

Here, we present an analysis of the spatio-temporal distribution of soil-transmitted helminth infection risks in Brazil. Our research extends a recent study that focused on the spatial distribution of soil-transmitted helminth infections in Brazil that was based on a relatively small database covering the period 2005–2009 (Scholte et al., 2013). We extended the survey period that now spans two decades (1995–2013) and focused on the space-time interactions of the disease patterns. We provide high-resolution spatial estimates of helminth species-specific infection risks and assess annualised deworming needs for school-aged children for Brazil. Historical data was extracted from the Global Neglected Tropical Diseases (GNTD) database (Hürlimann et al., 2011), and Bayesian spatio-temporal models were fitted in a SPDEs/INLA framework. Predictors included in each model were selected within a Bayesian geostatistical variable selection approach that is well suited for large datasets.

6.2 Methods

6.2.1 Disease data

Prevalence survey data pertaining to *A. lumbricoides*, hookworm and *T. trichiura* in Brazil were extracted from the GNTD database (www.gntd.org). The GNTD database is an open-access platform gathering spatially explicit survey data on soil-transmitted helminthiasis and other neglected tropical diseases identified through systematic searches of readily available electronic databases and grey literature (Hürlimann et al., 2011; Saarnak

et al., 2013). The literature search for relevant soil-transmitted helminth prevalence data in Brazil was updated on November 27 2013 and includes surveys conducted from 1995 onwards. The reader is referred to previous publications for further details on search strategy, geolocation and data quality appraisal (Hürlimann et al., 2011; Chammartin et al., 2013b).

Table 6.1: Data sources and properties of the predictors explored to model soil-transmitted helminth infection risk in Brazil.

Data type	Source	Temporal resolution	Spatial resolution
Temperature and precipitation[1]	WorldClim[2]	1950-2000	1 km
Altitude	SRTM[3]	2000	1 km
Soil acidity/soil moisture	ISRIC-WISE[4]	1960-2000	10 km
Human influence index (HII)	LTW[5]	2005	1 km
Human development index (HDI)	Atlas Brasil 2013[6]	2000 and 2010	Municipality
Poor households	Atlas Brasil 2013[6]	2000 and 2010	Municipality
Rurality, improved water supply, sewage system and waste treatment	Ministério da Saúde[7]	2010	Municipality
Population density	GPWFE[8]	2010	10 km

[1] A total of 19 climatic variables related to various factors were considered.

[2] WorldClim Global Climate database version 1.4. Available at: http://www.worldclim.org/ (accessed: March 2012).

[3] Shuttle Radar Topography Mission (SRTM). Available at: http://www.worldclim.org/ (accessed: March 2012).

[4] Global soil profile data ISRIC-WISE database version 1.2. Available at: http://www.isric.org/ (accessed: December 2012).

[5] Last of the Wild Data version 2, 2005 (LTW-2): Global Human Footprint Dataset (Geographic). Wildlife Conservation (WCS) and Center for International Earth Science Information Network (CIESIN). Available at: http://www.ciesin.org/wildareas/ (accessed: December 2013).

[6] Atlas do Desenvolvimento Humano no Brasil 2013. Available at: http://atlasbrasil.org.br/ (accessed: December 2013).

[7] Ministério da Saúde, Brazil. Available at: http://www2.datasus.gov.br/DATASUS/index.php (accessed: December 2013).

[8] Gridded Population of the World: future estimates (GPWFE): Center for International Earth Science Information Network (CIESIN), UN Food and Agriculture Organization (FAO) and Centro Internacional de Agricultura Tropical (CIAT). Available at: http://sedac.ciesin.columbia.edu/gpw (accessed: December 2012).

6.2.2 Environmental, socioeconomic and population data

Table 6.1 summarises the sources and the spatial and temporal resolutions of environmental, socioeconomic and population data considered in our analysis. A total of 29 variables were taken into account as potential risk factors for soil-transmitted helminth infection. Environmental data included altitude, soil acidity, soil moisture and 19 bioclimatic variables related to temperature and precipitation. Socioeconomic proxies were: human development index (HDI), which is a measure of socioeconomic development based on life expectancy, education and income; human influence index (HII), which quantifies human influence on ecosystems; a poverty measure reflected by the percentage of people living with a household monthly income lower than 60 US$ (poor households); and a measure of rurality expressed by the percentage of rural households within municipalities. In addition, using census data we compiled the proportion of individuals within municipalities with access to improved water supply, sewage system and waste treatment. These last three variables were classified as improved according to the following criteria: (i) sewage system connected to a network or to a septic tank; (ii) water supply from a well or through the network; and (iii) waste collection by a cleaning service.

Survey data were linked to potential risk factors based on their spatial proximity when they were available at fine spatial scale or according to their belonging to municipalities in case they were available at this resolution. Moreover, HDI and percentage of poor household data obtained in 2010 were assigned to prevalence data observed from 2005 onwards, while information obtained in 2000 was related to prevalence data prior to 2005.

6.2.3 Statistical analysis

Soil-transmitted helminth infection prevalence data were modelled via binomial logistic regression with spatio-temporal random effects accounting for a latent spatial process varying with time. Exploratory analyses were carried out to assess correlations between potential predictors, as well as to explore their association with observed infection risks. Highly correlated potential risk factors (Pearson's correlation coefficient > 0.9) were grouped, with the aim to include not more than one of them in the models. Continuous predictors were standardised (by substracting their mean and dividing with standard deviation) to obtain estimates of the effects, which are comparable across the predictors.

Details on spatio-temporal model formulation and variable selection are given in the Appendixes 6.6.1 and 6.6.2. In brief, risk factors included in the spatio-temporal models were selected through a Bayesian stochastic search variable selection approach (George

and McCulloch, 1993). We followed our previous procedure, which consists of selecting within a geostatistical framework the best predictors among highly correlated ones, while addressing non-linearity of the predictors Chammartin et al. (2013c). We further extended this formulation in applications to large datasets, by estimating the spatial process through a predictive process approximation (Banerjee et al., 2008). The inclusion of a variable in the model was defined as the product of two indicators: the first was assumed to be Bernoulli distributed and suggests the inclusion of the group of highly correlated variables, whilst the second followed a categorical prior distribution for selecting a single predictor within the group. In addition, regression coefficients were a priori parameterised with parameter expanded normal mixture of inverse-gamma (peNMIG) distributions (Scheipl et al., 2012), which ensure a rigorous selection of categorical variables. Models with the highest posterior probability identified the predictors to include in the final models.

Spatio-temporal distribution of soil-transmitted helminth infection risk was modelled using the methodology developed by Cameletti et al. (2013) for spatio-temporal modelling. SPDEs were used to represent a Matérn spatio-temporal Gaussian field (GF) as a Gaussian Markov random field (GMRF), which in turn allowed an INLA algorithm to estimate model parameters. This approach provides considerable advantages in terms of computational cost compared to traditional MCMC algorithms. The spatio-temporal GF is characterised by a first-order autoregressive temporal effect and another temporally independent effect assumed to arise from a zero mean multivariate normal distribution with spatio-temporal covariance function of the Matérn family for identical time periods.

We further predict the risk of infection with individual soil-transmitted helminth species over a grid of 381,881 pixels (5×5 km spatial resolution). To validate our models, we re-fitted our spatio-temporal models on a randomly selected subset of approximately 80% of the data, and compared model-based estimated risks with the remaining 20% observed prevalences. Model predictive ability was measured by the proportion of correctly predicted values within the kth highest posterior density (HPD) interval with k% probability coverage of the posterior distribution varying from 50% to 95%. We used the mean error (ME) to assess the prediction bias.

6.2.4 Population-adjusted risk and estimated treatment needs for school-aged children

The overall risk of soil-transmitted helminth infection was calculated for each of the samples of the predictive distribution, at each pixel, with a simple probabilistic model of

combined infection divided by a factor of 1.06 (De Silva and Hall, 2010). To calculate population-adjusted risks, we multiplied predicted risks by the population at pixel level, summed them up over areas of interest, and divided them by the population of those areas.

Annualised treatment needs for school-aged children (age range: 5–14 years) for preventive chemotherapy were estimated by considering one treatment per year for children living in low-risk municipalities (population-adjusted risk between 20 and 50%) and two treatments for children living in high-risk areas (population-adjusted risk $\geq$ 50%), following WHO guidelines (WHO, 2006). The school-aged population was estimated to represent 16.9% of the total population in Brazil, according to 2010 census data (http://www.ibge.gov.br/).

6.2.5 Ethics statement

All data were obtained from existing databases without personal identifiers. Here, the data were further analysed to deepen our understanding of the spatio-temporal distribution of soil-transmitted helminth infections in Brazil. Hence, there were no specific ethical considerations for the current analysis.

6.3 Results

From 1995 onwards, we obtained spatially explicit information about prevalence of *A. lumbricoides*, *T. trichiura* and hookworm across Brazil for 10,513, 10,497 and 10,492 locations, respectively. The frequency distribution of individual soil-transmitted helminth species surveys, stratified by year, is depicted in Figure 6.1. The datasets included 1,587, 1,572 and 1,570 unique locations for *A. lumbricoides*, *T. trichiura* and hookworm, respectively. Data were aggregated over four time periods, i. e. (i) 1995–1999; (ii) 2000–2004; (iii) 2005–2009; and (iv) from 2010 onwards. Figure 6.2 shows the spatial distribution of the observed prevalence, stratified by soil-transmitted helminth species and time periods. As illustrated in Figure 6.3, a reduction of the overall raw prevalence was observed over the four periods, with the exception of *T. trichiura* infection, which showed peak prevalence in 2000–2004.

A) *A. lumbricoides* B) *T. trichiura* C) Hookworm

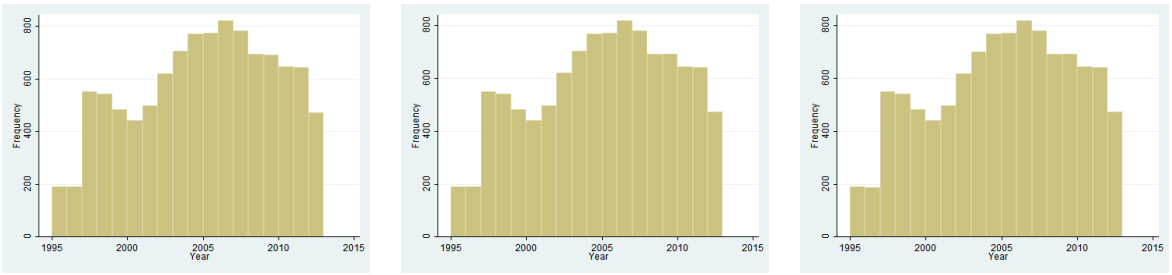


Figure 6.1: Frequency distribution of soil-transmitted helminth survey data in Brazil from 1995 to 2013, stratified by year. (A) *A. lumbricoides*, (B) *T. trichiura* and (C) hookworm.

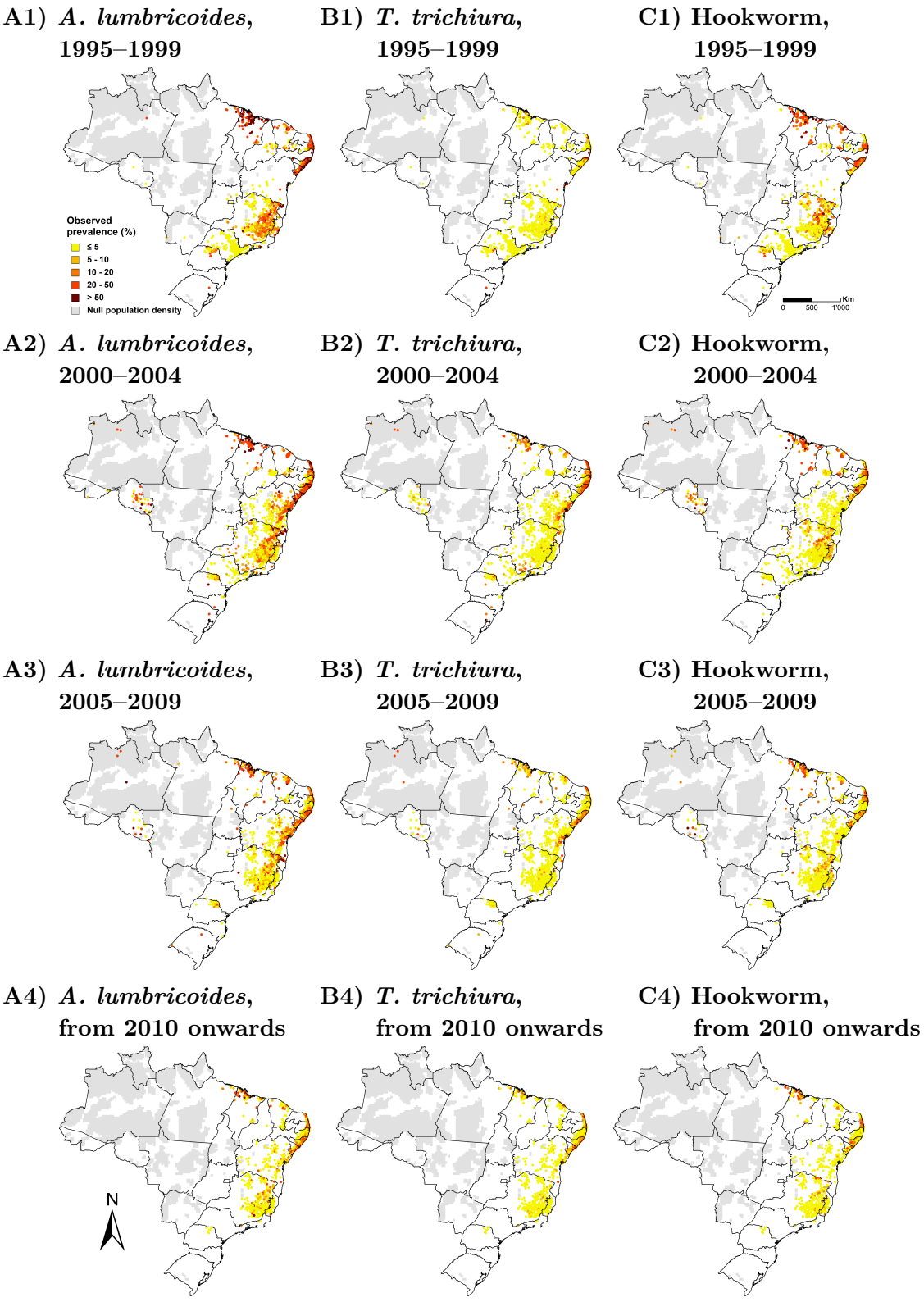


Figure 6.2: Observed soil-transmitted helminth prevalence in Brazil, stratified by species and 5-year time periods. (A) *A. lumbricoides*, (B) *T. trichiura* and (C) hookworm; (1) 1995–1999, (2) 2000–2004, (3) 2005–2009 and (4) from 2010 onwards.

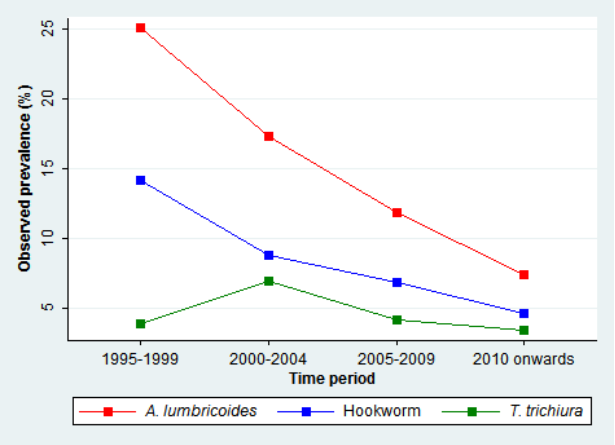


Figure 6.3: Temporal trend and observed national prevalences for *A. lumbricoides*, *T. trichiura* and hookworm infections in Brazil.

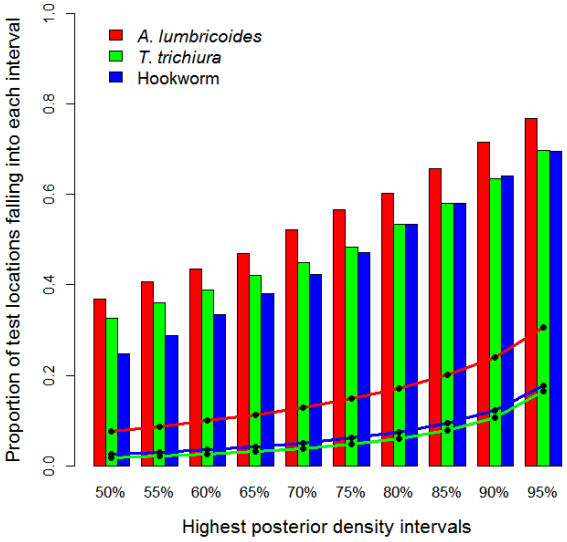


Figure 6.4: Model validation results. Proportion of surveys with prevalence of infection falling in the predicted highest posterior density (HPD) intervals (bar plots) for *A. lumbricoides*, *T. trichiura* and hookworm. The line plots show the corresponding width of the predicted HDP region.

Table 6.2: Variables selected by a Bayesian variable selection applied within the geostatistical logistic regression model.

	***A. lumbricoides* infection**	***T. trichiura* infection**	**Hookworm infection**
Group 1			
Yearly mean temperature[1]	0	0	0
Maximum temperature of warmest month[1]	0	0	0
Minimum temperature of coldest month[1]	0	0	0
Mean temperature of wettest quarter	0	0	0
Mean temperature of driest quarter	0	0	0
Mean temperature of warmest quarter[1]	X	0	X
Mean temperature of coldest quarter[1]	0	X	0
Group 2			
Mean diurnal temperature range[2]	X	0	0
Yearly temperature range[1,2]	0	X	X
Group 3			
Isothermality	X	X	0
Temperature seasonality	0	0	X
Group 4			
Yearly precipitation[1]	X	0	0
Precipitation in wettest month	0	0	0
Precipitation in wettest quarter[1]	0	X	X
Group 5			
Precipitation in driest month[1,3]	0	X	0
Precipitation in driest quarter[3]	X	0	X
Variables moderately correlated			
Precipitation seasonality	X	X	X
Precipitation in warmest quarter[1]	X	X	X
Precipitation in coldest quarter[1,3]	X	X	X
Altitude	X	0	X
Soil moisture[1,2,3]	X	X	X
Soil pH[1,3]	X	X	X
Human development index	X	X	X
Human influence index[1]	0	X	0
Rural households[1,3]	0	X	0
Improved sanitation	0	0	0
Improved water supply[1,2,3]	0	0	0
Improved waste collection[1]	0	0	0
Poor households	X	X	0
Survey period	fixed	fixed	fixed
Posterior probability	44.8%	93.5%	25.3%

X: selected; 0: not selected; [1]Categorised for *T. trichiura*; [2]Categorised for hookworm; [3]Categorised for *T. trichiura*.

The best model selected by the geostatistical variable selection is presented for each soil-transmitted helminth species, together with its posterior probability.

Table 6.3: Parameter estimates of bivariate and Bayesian spatio-temporal logistic models for *A. lumbricoides* infection risk in Brazil.

***A. lumbricoides* infection**	**Bivariate logistic†**		**Spatio-temporal model**	
	OR	**95% CI**	**OR**	**95% BCI**
Survey period				
1995–1999	1.00		1.00	
2000–2004	0.62	(0.56; 0.69)*	0.60	(0.51; 0.70)*
2005–2009	0.40	(0.34; 0.47)*	0.34	(0.28; 0.40)*
From 2010 onwards	0.24	(0.19; 0.30)*	0.14	(0.12; 0.17)*
Mean temperature of warmest quarter	1.91	(1.73; 2.12)*	0.98	(0.77; 1.24)
Mean diurnal temperature range	0.55	(0.50; 0.61)*	0.83	(0.73; 0.95)*
Isothermality	1.39	(1.25; 1.55)*	1.01	(0.90; 1.13)
Yearly precipitation	1.44	(1.32; 1.57)*	1.62	(1.43; 1.83)*
Precipitation in driest quarter (mm)				
< 50	1.00		1.00	
50–95	1.04	(0.75; 1.46)	1.56	(1.27; 1.92)*
≥ 95	1.83	(1.34; 2.51)*	1.36	(1.01; 1.84)*
Precipitation seasonality	0.88	(0.79; 0.97)*	0.92	(0.82; 1.02)
Precipitation in warmest quarter	0.53	(0.47; 0.59)*	0.65	(0.55; 0.76)*
Precipitation in coldest quarter (mm)				
< 80	1.00		1.00	
80–300	1.32	(0.98; 1.77)	0.65	(0.52; 0.81)*
≥ 300	4.21	(3.25; 5.45)*	0.92	(0.66; 1.30)
Altitude	0.49	(0.44; 0.55)*	0.82	(0.64; 1.06)
Soil moisture (%)				
< 50	1.00		1.00	
50–80	1.52	(1.21; 1.91)*	1.16	(0.96; 1.40)
≥ 80	0.92	(0.69; 1.22)*	1.01	(0.77; 1.33)
Soil pH				
< 5.35	1.00		1.00	
5.35–5.65	0.70	(0.54; 0.91)*	1.59	(1.36; 1.86)*
≥ 5.65	0.76	(0.59; 0.97)*	0.88	(0.74; 1.06)
Human development index (HDI)	0.62	(0.56; 0.69)*	1.29	(1.12; 1.49)*
Poor households	1.94	(0.68; 2.25)	1.81	(1.52; 2.15)*
			Median	**95% BCI**
Temporal autocorrelation			0.03	(-0.02; 0.07)
Spatial variance			5.07	(4.73; 5.31)
Spatial range (km)			30.2	(28.1; 35.2)

* Significant based on 95% CI or 95% BCI.

† With standard error clustered at location level.

OR: odds ratio; 95% CI: lower and upper bound of a 95% confidence interval; 95% BCI: lower and upper bound of a 95% Bayesian credible interval.

Table 6.4: Parameter estimates of bivariate and Bayesian spatio-temporal logistic models for *T. trichiura* infection risk in Brazil.

***T. trichiura* infection**	**Bivariate logistic†**		**Spatio-temporal model**	
	OR	**95% CI**	**OR**	**95% BCI**
Survey period				
1995–1999	1.00		1.00	
2000–2004	1.83	(1.50; 2.23)*	3.47	(2.74; 4.40)*
2005–2009	1.05	(0.82; 1.36)	1.66	(1.30; 2.10)*
From 2010 onwards	0.87	(0.64; 1.17)	0.93	(0.71; 1.21)
Yearly temperature range (°C)				
< 13	1.00		1.00	
13–18	0.19	(0.15; 0.25)*	0.34	(0.25; 0.47)*
≥ 18	0.12	(0.09; 0.17)*	0.34	(0.22; 0.53)*
Mean temperature of coldest quarter (°C)				
< 19	1.00		1.00	
19–22	3.03	(2.25; 4.07)*	0.83	(0.61; 1.13)
≥ 22	7.97	(6.09; 10.42)*	1.42	(0.92; 2.18)
Isothermality	1.28	(1.15; 1.42)*	1.27	(1.09; 1.48)*
Precipitation seasonality	0.70	(0.64; 0.76)*	0.82	(0.68; 0.98)*
Precipitation in wettest quarter (mm)				
< 560	1.00		1.00	
560–680	1.03	(0.74; 1.45)	1.65	(1.22; 2.24)*
≥ 680	1.72	(1.24; 2.40)*	2.11	(1.48; 3.00)*
Precipitation in warmest quarter (mm)				
< 250	1.00		1.00	
250–440	0.62	(0.47; 0.83)*	1.50	(1.14; 1.99)*
≥ 440	0.18	(0.13; 0.24)*	2.45	(1.46; 4.11)*
Precipitation in driest month (mm)				
< 14	1.00		1.00	
14–16	1.07	(0.70; 1.64)	1.48	(1.10; 1.97)*
≥ 16	3.21	(2.16; 4.79)*	2.00	(1.32; 3.05)*
Precipitation in coldest quarter (mm)				
< 80	1.00		1.00	
80–300	2.04	(1.38; 3.02)*	1.13	(0.82; 1.55)
≥ 300	7.67	(5.48; 10.75)*	1.80	(1.11; 2.92)*
Soil moisture (%)				
< 50	1.00		1.00	
50–80	1.56	(1.15; 2.13)*	0.82	(0.61; 1.09)
≥ 80	0.91	(0.54; 1.52)	0.57	(0.38; 0.86)*
Soil pH				
< 5.35	1.00		1.00	
5.35–5.65	0.54	(0.38; 0.77)*	1.52	(1.20; 1.92)*
≥ 5.65	0.60	(0.45; 0.80)*	0.87	(0.67; 1.12)

Continued on next page

***T. trichiura* infection**	**Bivariate logistic†**		**Spatio-temporal model**	
	OR	**95% CI**	**OR**	**95% BCI**
Human development index (HDI)	0.74	(0.65; 0.84)*	1.45	(1.17; 1.79)*
Human influence index (HII)				
< 20	1.00		1.00	
20–26	1.39	(1.01; 1.91)*	1.49	(1.20; 1.86)*
≥ 26	2.14	(1.57; 2.91)*	1.86	(1.46; 2.37)*
Rural households (%)				
< 25	1.00		1.00	
25–50	1.01	(0.74; 1.39)	0.86	(0.68; 1.09)
≥ 50	0.61	(0.44; 0.74)*	0.66	(0.51; 0.85)*
Poor households	1.54	(1.27; 1.87)*	2.06	(1.58; 2.68)*
			Median	**95% BCI**
Temporal autocorrelation			-0.05	(-0.10; 0.00)
Spatial variance			9.68	(9.27; 10.03)
Spatial range (km)			32.2	(29.9; 33.9)

* Significant based on 95% CI or 95% BCI.

† With standard error clustered at location level.

OR: odds ratio; 95% CI: lower and upper bound of a 95% confidence interval; 95% BCI: lower and upper bound of a 95% Bayesian credible interval.

Table 6.5: Parameter estimates of bivariate and Bayesian spatio-temporal logistic models for hookworm infection risk in Brazil.

Hookworm infection	**Bivariate logistic**[†]		**Spatio-temporal model**	
	OR	**95% CI**	**OR**	**95% BCI**
Survey period				
1995–1999	1.00		1.00	
2000–2004	0.58	(0.50; 0.68)*	0.54	(0.43; 0.68)*
2005–2009	0.44	(0.36; 0.54)*	0.28	(0.22; 0.35)*
From 2010 onwards	0.29	(0.22; 0.39)*	0.13	(0.10; 0.17)*
Mean temperature of warmest quarter	2.00	(1.68; 2.37)*	1.50	(1.10; 2.05)*
Temperature seasonality	0.54	(0.43; 0.67)*	1.48	(1.21; 1.81)*
Yearly temperature range (°C)				
< 13	1.00		1.00	
13–18	0.61	(0.44; 0.86)*	0.98	(0.69; 1.37)
≥ 18	0.20	(0.15; 0.27)*	0.94	(0.60; 1.48)
Precipitation in coldest quarter	1.62	(1.43; 1.85)*	0.89	(0.69; 1.15)
Precipitation in warmest quarter	0.44	(0.35; 0.55)*	0.43	(0.33; 0.55)*
Precipitation seasonality	1.29	(1.15; 1.45)*	0.61	(0.49; 0.77)*
Precipitation in driest quarter	0.88	(0.76; 1.01)	0.74	(0.62; 0.87)*
Precipitation in wettest quarter	1.54	(1.31; 1.80)*	3.59	(2.79; 4.62)*
Altitude	0.56	(0.48; 0.67)*	0.99	(0.71; 1.37)
Soil moisture (%)				
< 50	1.00		1.00	
50–80	0.89	(0.64; 1.23)	0.52	(0.40; 0.67)*
≥ 80	0.49	(0.31; 0.78)*	0.30	(0.20; 0.43)*
Soil pH	0.87	(0.76; 1.00)	0.77	(0.70; 0.86)*
Human development index (HDI)	0.58	(0.53; 0.63)*	0.74	(0.68; 0.82)*
			Median	**95% BCI**
Temporal autocorrelation			0.01	(-0.07; 0.06)
Spatial variance			8.92	(8.42; 9.43)
Spatial range (km)			29.7	(28.0; 31.4)

* Significant based on 95% CI or 95% BCI.

† With standard error clustered at location level.

OR: odds ratio; 95% CI: lower and upper bound of a 95% confidence interval; 95% BCI: lower and upper bound of a 95% Bayesian credible interval.

Results of the variable selection are given in Table 6.2. Out of the 29 potential predictors investigated, our variable selection procedure identified 14, 13 and 12 variables as being important for *T. trichiura*, *A. lumbricoides* and hookworm, respectively with model posterior probabilities of 93.5%, 44.8% and 25.3%. The selected variables were subsequently used to build spatio-temporal models.

Parameter estimates of spatio-temporal multiple regression models, together with the ones of bivariate logistic associations with standard error clustered at location-level are presented for each soil-transmitted helminth species in Tables 6.3–6.5. Results of bivariate logistic regressions show associations of the selected predictors with observed risk. Temperature and precipitation usually favour the risk of soil-transmitted helminthiasis, as reflected by the positive bivariate associations of temperature during warmest and coldest quarters and precipitation in coldest quarter and coldest month. However, precipitation during the warmest quarter was negatively associated with the risk of infection with any of the three soil-transmitted helminth species. Furthermore, important temperature and precipitation oscillations show a negative association with soil-transmitted helminth infection odds, as suggested by the effects of diurnal and yearly temperature ranges, low isothermality, as well as temperature and precipitation seasonality. The three infection risks were positively associated to proxies of poverty, as reflected by the positive effect of the percentage of poor households and the negative association of HDI.

In the spatio-temporal model, the odds of *A. lumbricoides* infection risk was positively associated with yearly precipitation, precipitation in driest quarter, soil pH (5.35–5.65), poor households and HDI, and negatively associated with mean diurnal temperature range, precipitation in warmest quarter and coldest quarter (80–300 mm).

For *T. trichiura*, the predictors with an important positive effect on the odds of the risk were: isothermality, precipitation in driest month, wettest, warmest and coldest quarters, soil pH (5.35–5.65), HDI, HII and poor households. On the other hand, the odds of *T. trichiura* infection were negatively associated with yearly temperature range, precipitation seasonality, soil moisture ($\geq 80\%$) and rural households.

Hookworm infection odds increased with average temperature of warmest month, temperature seasonality, as well as with precipitation in the wettest quarter. On the other hand, precipitation in the warmest and driest quarter, precipitation seasonality, soil moisture, pH and HDI were negatively associated with the risk of hookworm infection.

The estimates of the effects of the survey periods indicate a decreasing trend for both *A. lumbricoides* and hookworm infection risks in the period studied, i. e. from 1995 onwards

until late 2013. For *T. trichiura*, there was no important effect of survey period after 2010 compared to the preceding decade.

Figure 6.4 shows the results of model validation. The risks of soil-transmitted helminth infection were correctly predicted within a 95% credible interval for 77% of the tested data for *A. lumbricoides*, 70% for *T. trichiura* and 69% for hookworm. The ME was −3.03%, −2.26% and −2.75% for the three species, respectively, suggesting that our models slightly underestimate the observed prevalences.

Model-based predictions of the geographical distribution of the three soil-transmitted helminth species considered in our analyses are presented in Figure 6.5, for each of the four time periods. From 2010 onwards, *A. lumbricoides* infection presents larger risk areas compared to the other two species, with higher risk in the northern part of the country. The highest risk for *T. trichiura* was found in the north-western part of Brazil, while the risk for hookworm was higher along the northern coast. Our maps also highlight the temporal evolution of the risk for an infection with any of these three soil-transmitted helminth species over the past 20 years. Apparent shrinkage of high-risk areas was observed for *A. lumbricoides* and hookworm. Spatial correlation was estimated around 30 km for each of the three soil-transmitted helminth species, and spatial variance extended from 5.07 to 9.68. Temporal autocorrelation was generally weak, suggesting that temporal structure was explained by the temporal trend, as well as by changes in HDI and percentage of poor households over time.

Predicted population-adjusted risk estimates in Brazil are given for each survey period analysed (Table 6.6). Based on predictions from 2010 onwards, we estimated that 10.9 million people were infected with soil-transmitted helminths in Brazil (population-adjusted risk = 6.0%; 95% Bayesian credible interval (BCI): 5.4–6.9%). Single species infection population-adjusted risks were estimated at 3.6% for *A. lumbricoides* (95% BCI: 3.0–4.3%), 1.7% for hookworm (95% BCI: 1.4-2.3%), and 1.4% for *T. trichiura* (95% BCI: 1.1–1.7%). Low-risk (population-adjusted risk 20–50%) and high-risk (population-adjusted risk $\geq$ 50%) municipalities are highlighted in Figure 6.6. The highest population-adjusted risk of soil-transmitted helminthiasis was found along the northern coast. We estimated that 1.8 million doses of anthelminthic treatments are required each year for preventive chemotherapy targeting school-aged children at municipality level in Brazil.

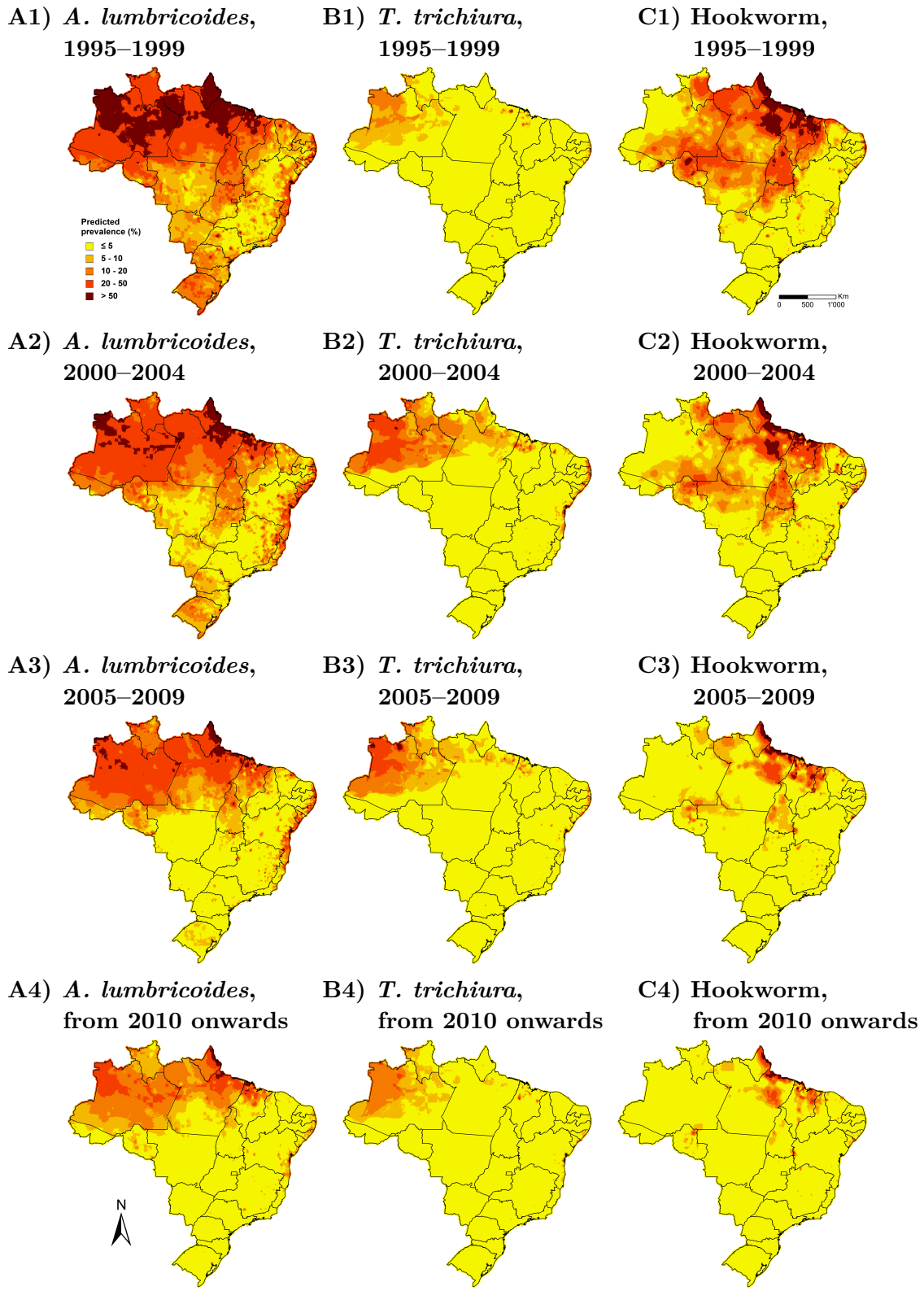


Figure 6.5: Predicted soil-transmitted helminth risk in Brazil, stratified by species and 5-year time periods. (A) *A. lumbricoides*, (B) *T. trichiura* and (C) hookworm; (1) 1995–1999, (2) 2000–2004, (3) 2005–2009 and (4) from 2010 onwards.

Table 6.6: Predicted population-adjusted risk of *A. lumbricoides*, *T. trichiura*, hookworm and overall soil-transmitted helminth infection in Brazil, stratified by survey period.

Survey period	*A. lumbricoides* infection risk (%)	*T. trichiura* infection risk (%)	Hookworm infection risk (%)	Soil-transmitted helminth infection risk (%)
1995–1999	15.6 (13.6; 18.0)	1.8 (1.4; 2.3)	7.6 (6.6; 9.0)	20.9 (19.0; 23.2)
2000–2004	11.4 (10.0; 13.0)	4.5 (3.6; 5.7)	5.7 (4.9; 6.9)	17.9 (16.5; 19.7)
2005–2009	7.9 (6.8; 9.1)	2.5 (2.1; 3.2)	2.8 (2.4; 3.4)	11.5 (10.4; 12.6)
From 2010 onwards	3.6 (3.0; 4.3)	1.4 (1.1; 1.7)	1.7 (1.4; 2.3)	6.0 (5.4; 6.9)

Population-adjusted risks are given with their 95% Bayesian credible interval (BCI).
Risk is adjusted on population of 2010 for survey periods from 2005 onwards and on population of 2000 for survey periods prior to 2005.

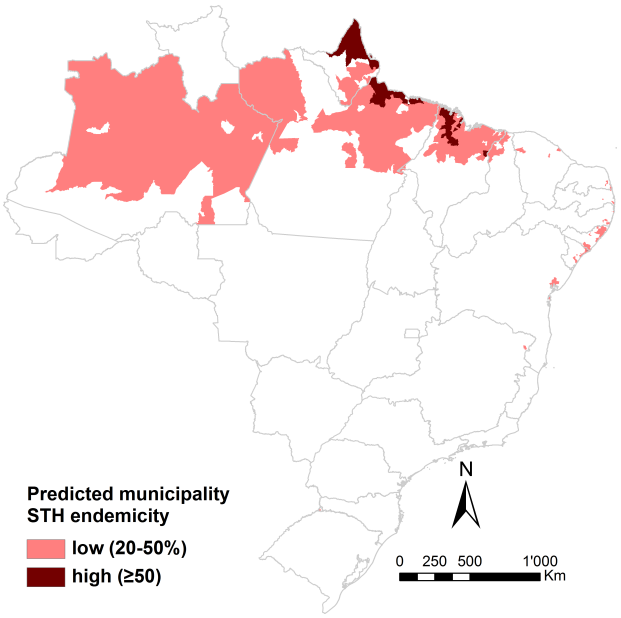


Figure 6.6: Estimated soil-transmitted helminthiasis endemicity of the Brazilian municipalities for intervention planning according to WHO guidelines pertaining to preventive chemotherapy.

6.4 Discussion

The current study focuses on the spatio-temporal distribution of *A. lumbricoides*, hookworm and *T. trichiura* risk in Brazil and therefore complements and expands upon a recent study that investigated spatial patterns (Scholte et al., 2013). We present predictive risk maps at high spatial resolution from 1995 onwards at 5-year increments. Additionally, we provide a map that highlights municipalities that require preventive chemotherapy targeting school-aged children according to recommendations put forward by WHO (2012). Our analyses provide new insight into spatio-temporal risk profiling of helminthiasis based on a large ensemble of geolocated survey data by considering space-time interactions.

We provide model-based evidence of a decrease of *A. lumbricoides* and hookworm infection risks over the past 20 years in Brazil. Interestingly, the temporal evolution of the third common soil-transmitted helminth species — *T. trichiura* — increased after 2000, started to decline from 2005 onwards, and finally reached similar levels to the situation in 1995–1999 from 2010 onwards. We believe that the main reasons explaining the lower risk of textitA. lumbricoides and hookworm from 2010 onwards compared to the situation 20 years ago are the social and economic development, coupled with deworming activities. Nevertheless, it is important to note that no mass deworming activities were carried out by the Ministry of Health (MoH) in Brazil from 2005 to 2011 (WHO, 2011b, 2012). The question arises why a similar decline was not observed for *T. trichiura*. Differences may reflect differential efficacies of the widely used deworming drugs albendazole and mebendazole. While both drugs show high cure and egg reduction rates against *A. lumbricoides*, and albendazole shows satisfactory efficacy against hookworm, neither drug results in high efficacy against *T. trichiura* (Keiser and Utzinger, 2008, 2010). These differences might explain the delayed risk change profile for *T. trichiura*.

Our predictive risk maps highlight that high-risk areas of *A. lumbricoides* and *T. trichiura* infections occur in the north-western part and along the eastern coast of Brazil, while high-risk areas of hookworm infection is concentrated along the northern coast. This is coherent with patterns highlighted by two previous analyses (Chammartin et al., 2013b; Scholte et al., 2013). However, our population-adjusted estimates for the period 2005–2009 of 7.9% for *A. lumbricoides*, 2.5% for *T. trichiura* and 2.8% for hookworm are smaller than those based on predictions from 2005 onwards stemming from a spatial analysis of South America (i. e. 14.3% for *A. lumbricoides*, 10.1% for *T. trichiura* and 12.3% for hookworm) (Chammartin et al., 2013b). These differences might be explained by the inclusion of socioeconomic factors in the current analysis. Our previous work did not include poverty

indicators due to the difficulty of deriving consistent measures among different countries. In comparison to a temporal trend included as a covariate, which indicates the change of magnitude of the risk over time (Chammartin et al., 2013b; Lai et al., 2013), spatio-temporal models, as developed in this analysis, highlight the changes in the geographical patterns of risk over time. Hence, our analysis highlights the importance of considering the temporal aspect of infection risk, especially in a country like Brazil, where socioeconomic conditions have considerably improved and infectious diseases risk has declined over time (Oliveira, 2010; Chammartin et al., 2013b). In comparison to Scholte et al. (2013), who analysed a restricted dataset with data provided by the national schistosomiasis control programme for the period 2005–2009, we estimated considerably smaller risks for both *A. lumbricoides* (7.9% *versus* 15.6%) and textitT. trichiura (2.5% *versus* 10.1%). We explain these differences by a considerably higher spatial coverage of our data. Recently, Pullan et al. (2014) estimated the global risk of soil-transmitted helminth infections for the year 2010 based on empirical approaches, which do not account for small-scale spatial variation. For Brazil, they estimated a risk between 1% and 10% for *T. trichiura* and hookworm infections, which is comparable to our estimates of 1.4% and 1.7% from 2010 onwards. However, we estimated a risk of 3.6% for *A. lumbricoides*, whilst Pullan et al. (2014) estimates a risk between 10% and 20%. This difference highlights the importance of capturing small-scale variation in estimating the risk of helminth infection and other neglected tropical diseases.

Parameter estimates of the spatio-temporal models reflect the climatic suitability and socioeconomic conditions that favour soil-transmitted helminthiasis transmission in Brazil. Each soil-transmitted helminth species risk is influenced by complex interactions of the predictors selected by our variable selection approach. In particular, our analysis confirms that warm and humid conditions are suitable for soil-transmitted helminth egg and larval development (Spindler, 1929; Brooker et al., 2006). Positive associations of precipitation were observed for the three soil-transmitted helminth species and temperature was an important risk factor for hookworm. Extreme weather conditions could adversely affect development and survival of helminthic free-living stages. Indeed, larvae optimally hatch within certain temperature limits (Brooker et al., 2004b), suggesting that extreme temperatures might impair their development. Larger range of temperature during the day showed a negative effect in *T. trichiura* and *A. lumbricoides* models, while strong isothermality positively impacts the risk of *T. trichiura*, confirming this hypothesis. Furthermore, it has been speculated that heavy rainfall might wash out soil-transmitted helminth eggs from the soil

(Brown, 1927; Gunawardena et al., 2004; Chammartin et al., 2013b). The negative effects of precipitation seasonality in *T. trichiura* and hookworm models, precipitation in warmest quarter in *A. lumbricoides* model, as well as of soil moisture in hookworm model point in that direction. We also note that high isothermality, low range of temperature during the day and low precipitation seasonality are typical characteristics of northern equatorial and tropical humid regions of Brazil, suggesting that those climatic areas are suitable for transmission. The optimal soil acidity for *A. lumbricoides* and *T. trichiura* transmission ranges between pH values of 5.35 and 5.65, however, hookworm prefers somewhat less acid conditions.

Our analysis also highlights the intimate connection of soil-transmitted helminth infection with poverty. Indeed, high percentage of poor households was an important risk factor for both *T. trichiura* and *A. lumbricoides* infections, after accounting for HDI. Poor households generally show lower frequencies of access and use of clean water and improved sanitation and thus are at higher odds of soil-transmitted helminth infection (Ziegelbauer et al., 2012; Strunz et al., 2014). Another interesting aspect is the positive effect of HII and low percentage of rural households associated with *T. trichiura* risk, confirming previous findings (Pullan and Brooker, 2012; Chammartin et al., 2013b). These observations suggest that *T. trichiura* infection might be more prevalent in urban compared to rural settings (Pullan and Brooker, 2012).

Most of the soil-transmitted helminth data (97.5%) stem from the Brazilian schistosomiasis control programme, which took advantage of the Kato-Katz technique, which allows concurrent diagnosis of soil-transmitted helminths, while screening for *Schistosoma mansoni* eggs in faecal thick smears. Brazil launched its national schistosomiasis control programme in 1975 with the aim of reducing schistosomiasis-related morbidity. Regarding soil-transmitted helminthiasis, the MoH re-started a mass deworming campaign for school-aged children in 2013, prioritising areas characterised by a low HDI. This campaign will now be extended to the whole of Brazil. Data generated by this programme will facilitate the study of the evolution of the risks and evaluation of the impact of interventions. It will be important to specifically inform about the situation regarding the population targeted by the interventions (e. g., school-aged children or entire communities). Data that we analysed in the current study were mainly collected within the whole population (only 1% referred to children exclusively). Thus, we might underestimate the risk among children, as it is known that they are usually at a higher risk of soil-transmitted helminth infections, particularly *A. lumbricoides* and *T. trichiura* (Anderson and May, 1985). Importantly

though, despite low HDI showing a positive association with the three soil-transmitted helminth infection risks in bivariate associations, our spatio-temporal analysis indicates that hel-minthiasis risk was driven by complex environmental/socioeconomic interactions. Hence, we believe that our estimates provide useful information for a refined target of interventions.

From a modelling point of view, it is important to highlight that SPDE methodology and INLA enabled us to fit our spatio-temporal models at a reasonable computational cost (around 4 hours per model, including fitting and prediction). Implementing this type of model would have been difficult to achieve with MCMC, especially because of convergence problems and large number of locations for spatio-temporal process estimation. With regard to the risk of the three soil-transmitted helminth species, residual spatial correlation was moderate (around 30 km) and temporal autocorrelation was weak. Hence, most of the spatial and temporal dynamics were captured by the covariates of our models.

6.5 Conclusions

The methodology employed in the current analysis enables fitting more complex models and provides a useful tool for joint analysis of space and time components for risk profiling. The analysis of the spatio-temporal aspect of the risk of soil-transmitted helminth infections deepens our understanding of the evolution of the risk across time and enables more accurate predictions of the infection risks. We hope that our estimates will provide useful benchmark information for the soil-transmitted helminthiasis control programme in Brazil to prioritise interventions and enhance spatial targeting.

Acknowledgements

The authors are grateful for financial support from the Swiss Brazilian Joint Research Programme (BSJRP 011008).

6.6 Appendix

6.6.1 Geostatistical variable selection formulation

Let Y_i, n_i and p_i be the number of infected individuals, the number of individuals screened, and the prevalence of infection at location i ($i = 1, ..., N$). We assume that Y_i arises from a Binomial distribution, i. e. $Y_i \sim Bn(n_i, p_i)$. Potential predictors $X_{j_b}^{(b)}$, $j_b = 1, ..., J_b$ are divided into B groups b, ($b = 1, ..., B$), where $B - 1$ groups contain predictors wich are considered highly correlated with a Pearson coefficient > 0.9, while the Bth group includes predictors that exhibit only moderate correlation with other potential predictors. In addition, potential predictors presenting a non-linear association to the infection risk in explanatory analyses have been categorised and we define $X_{lj_b}^{(b)}$ as being the lth categorical form of predictor $X_{j_b}^{(b)}$, where $l = 1, ..., L$ categories (excluding baseline). Our variable selection procedure aims to select most important predictors, while accounting for spatial correlation, addressing non-linearity of the predictors, and forcing the model to choose a maximum of one predictor among the ones considered as highly correlated. To that end, we model a categorical temporal trend T_{il} ($l = 1, ..., L$ categories), the potential predictors $X_{lj_bi}^{(b)}$ and a spatial random effect φ_i on the logit scale, such as:

$$logit(p_i) = \beta_0 + \sum_{l=1}^{L} T_{il}\beta_{1l} + \sum_{b=1}^{B}\sum_{j_b=1}^{J_b} \alpha_{j_b} \sum_{l=1}^{L_{j_b}} \xi_{lj_b} X_{lj_bi}^{(b)} + \varphi_i,$$

where regression coefficients of potential predictors X_{j_b} are defined as the product of an overall contribution α_{j_b} and the effect ξ_{lj_b} of each of its elements (i. e. categories).

Within a Bayesian framework of inference, we assign a spike and slab prior (Scheipl et al., 2012; Chammartin et al., 2013a,c) to α_{j_b}, which is a scaled normal mixture of inverse-gamma, that is $\alpha_{j_b} \sim N(0, \tau_{j_b}^2)$, where $\tau_{j_b}^2 \sim \gamma_1^{(b)}\gamma_{2j_b}^{(b)} IG(a_\tau, b_\tau) + (1 - \gamma_1^{(b)}\gamma_{2j_b}^{(b)})\upsilon_0 IG(a_\tau, b_\tau)$, where a_τ and b_τ are fixed parameters of non-informative inverse-gamma distribution set to 5 and 25, respectively, while υ_0 is a small constant set to 0.00025, shrinking α_{j_b} to zero when the predictor is excluded. The product of the two indicators $\gamma_1^{(b)}$ and $\underline{\gamma}_2^{(b)} = (\gamma_{21}^{(b)}, ..., \gamma_{2J_b}^{(b)})^T$ indicates the presence or absence of the predictors in the model. In particular, $\gamma_1^{(b)}$ determines the presence or absence of the group b in the model and $\gamma_{2j_b}^{(b)}$ allows selection of a single predictor within the group. A Bernoulli and a categorical prior distribution are assigned to $\gamma_1^{(b)}$ and $\underline{\gamma}_2^{(b)}$, respectively, such as $\gamma_1^{(b)} \sim Bern(\Omega_1^{(b)})$ and $\underline{\gamma}_2^{(b)} \sim Cat(J_b, \Omega_{21}^{(b)}, ..., \Omega_{2J_b}^{(b)})$ with inclusion probabilities $\Omega_1^{(b)}$ and $\underline{\Omega}_2^{(b)}$. To allow greater flexibility in estimating model size, these probabilities are considered as hyper-parameters having non-informative beta and Dirichlet distributions; $\Omega_1^{(b)} \sim Beta(1, 1)$, $\underline{\Omega}_2^{(b)} = (\Omega_{21}^{(b)}, ..., \Omega_{2J_b}^{(b)})^T \sim Dirichlet(1, ..., 1)$. A mixture of two Gaussian distributions is

assumed for ξ_{lj_b}, $\xi_{lj_b} \sim N(m_{lj_b}, 1)$, $m_{lj_b} \sim 1/2\delta_1(m_{lj_b}) + 1/2\delta_{-1}(m_{lj_b})$, which shrinks ξ_{lj_b} towards $|1|$ (multiplicative identity). For predictors moderately correlated, γ_{2j_b} is fixed to 1, while the effect of linear predictors is only defined by an overall contribution of α. In addition, non-informative normal priors have been assigned to the constant β_0 and the effects β_{1l} of the temporal trend; $\beta_0, \beta_{1l} \sim N(0, 100)$.

Large matrix computation cost in estimating this latent spatial process φ is overcome with the predictive process estimation (Banerjee et al., 2008). In more details, φ is estimated from a subset of 200 locations (knots) $\{s_k^*, k = 1, .., K\}$ with latent observations $\varphi^* = (\varphi_1^*, ..., \varphi_K^*)^T$, $\varphi^* \sim MVN(0, \Sigma^*)$. Σ^* is the $K \times K$ variance-covariance matrix modelled by an isotropic exponential correlation function of distance, i. e, $\Sigma_{cd}^* = \sigma_{sp}^2 exp(-\rho d_{cd})$, where d_{cd} is the Euclidean distance between locations c and d, σ_{sp}^2 is the geographical variability, and ρ controls the rate of correlation decay. Inverse gamma distribution $\sigma_{sp}^2 \sim IG(2.01, 1.01)$ is chosen for the variance σ_{sp}^2 and a gamma distribution is assumed for the spatial decay ρ, $\rho \sim G(0.01, 0.01)$. Spatial random effect φ at original set of locations are predicted via the conditional mean $Q^T \Sigma^{*-1} \varphi^*$, where $Q = Cov(\varphi^*, \varphi)$ is a $N \times K$ matrix of the covariance function between the K knots and the N observed locations. Minimax space filling sampling (Johnson et al., 1990; Diggle and Lophaven, 2006) is used to select the knots using the cover.design routine in R (R Core Team, 2014).

Geostatistical variable selection was run in JAGS through the rjags library of R (R Core Team, 2014) in JAGS 3.4.0 with one chain sampler and 40,000 iterations (including a burn-in of 10,000 iterations). Final 10,000 iterations were used to calculate models posterior probabilities and the subset of variables included in the models with the highest posterior probabilities identified the final models.

6.6.2 Bayesian spatio-temporal model formulation

Our Bayesian spatio-temporal model formulation follows the approach introduced by Cameletti et al. (2013). In particular, we define Y_{it}, p_{it} and n_{it} as the number of infected individuals, the number of individuals screened, and the prevalence of infection at location i $(i = 1, ..., N)$ for time t $(t = 1, ..., T)$, and we assume Y_{it} to be generated from a binomial distribution, i. e. $Y_{it} \sim Bin(p_{it}, n_{it})$. Prevalence of infection is then linearly regressed on the logit scale as follows: $logit(p_{it}) = X_{it}^T \underline{\beta} + \varphi_{it}$, where X is the matrix of explanatory variables (including an intercept, a temporal trend, and the predictors selected by the variable selection), $\underline{\beta}$ is the regression coefficient vector, and $\underline{\varphi}$ is a spatio-temporally-structured random effect. We allow the spatio-temporal process φ_{it} to change in time with

a first order autoregressive process (AR1), such as:

$$\begin{aligned} \varphi_{it} &= \omega_{i1} && \text{if } t = 1 \\ \varphi_{i,t} &= a\varphi_{i,t-1} + \omega_{i,t} && \text{if } t = 2, .., T, \end{aligned}$$

with a temporal autoregressive coefficient a, $|a| < 1$ and a temporally independent spatially-structured effect ω which is assumed to be multivariate normal with zero mean and spatio-temporal covariance function of the Matérn family:

$$Cov(\omega_{i,t}, \omega_{j,t'}) = \begin{cases} 0 & \text{if } t \neq t' \\ \sigma^2_\omega C(d_{ij}) & \text{if } t = t' \end{cases} for\, i \neq j,$$

where σ^2_ω is the variance of the structured effect ω, (σ^2_ω=Var($\omega_{i,t}$)). The spatial correlation function $C(d_{ij})$ is function of the Euclidean distance between locations i and j (d_{ij}) and is defined by the Matérn function given by:

$$C(d_{ij}) = \tfrac{1}{\Gamma(\upsilon)2^{\upsilon-1}}(\kappa d_{ij})^\upsilon K_\upsilon(\kappa d_{ij}),$$

where K_υ is the modified Bessel function of second kind and order υ ($\upsilon > 0$), υ is a smoothing parameter controlling the rate of correlation decay fixed to 1, and κ ($\kappa > 0$) is a scaling parameter. The spatial range is defined as the minimum distance at which spatial correlation between locations is less than 10% and is given by $\sqrt{8\upsilon/\kappa}$.

Chapter 7

Discussion and outlook

The work presented in this PhD thesis contributes to the field of epidemiology of helminths and develops Bayesian statistical methodology for spatial modelling and estimation of infection risks. Our work takes place in a context where control is scaling up and reliable spatially explicit estimates of the risk are needed for cost-effective guidance and planning of interventions. This PhD thesis is structured following a journal-article format, and includes five manuscripts where methodology, results and limitations are discussed in details. This section aims to synthesise the significance of our work by discussing our main contributions and opens further extensions.

7.1 Significance of the work

7.1.1 Contribution in spatial modelling of helminthiases risk

The selection of the best set of predictors is an essential step in statistical modelling. Environment is known to play an important role for the development and survival of the free living stages of helminths, while poor socioeconomic conditions favour infection transmission. Various socio-environmental proxies are available from diverse sources, and helminthiases risk can be explained by complex interactions of a set of risk factors, depending on the setting. In Chapter 2, we addressed limitations of traditional variable selection procedures. Variable selection within a geostatistical framework has captured little attention. Most exercises ignore the spatial correlation, which can lead to incorrect estimates of the statistical significance of the predictors included in the model. We have addressed this issue by implementing different parameterisations of the stochastic search variable selection (George and McCulloch, 1993). We first included "parameter expanded normal mixture of inverse-gamma" prior distributions (Scheipl et al., 2012) for regression coefficients, which allow rigorous selection of blocks of covariates, particularly categorical variables (Chapter 2). Potential predictors can also be highly correlated. Ignoring this aspect can lead to conflicting effects adding unnecessary noise. Hence, we further developed our approach by introducing an indicator *a priori* parameterised with a categorical distribution which suggests the inclusion of at most one predictor among the highly correlated ones (Chapter 5). We then extended our selection strategy to applications on large datasets, where the spatial process was estimated from a subset of locations (Chapter 6), as well as to a joint modelling approach of two parasite species, where a multinomial model was used to analyse the different infection status, including co-infection (Chapter 3).

Estimation of very large number of model parameters is a computational challenge and Markov chain Monte Carlo (MCMC) algorithms are known to converge slowly. When the

number of locations is large (> 1000), it becomes difficult to estimate the spatial process without relying on low-rank approximations. In Chapter 4, we used the predictive process approximation (Banerjee et al., 2008) to model large datasets and estimated spatially-structured random effects from a subset of locations. We have also overcome MCMC simulations in Chapter 6 by using integrated nested Laplace approximation (INLA) (Rue et al., 2009) and approximated the Gaussian spatial process by a Gaussian Markov random field (GMRF) through stochastic partial differential equations (SPDE) (Lindgren et al., 2011). The Markovian properties of the GMRF involve sparsity of the spatial process precision matrix, which allows INLA to estimate model parameters of complex and large datasets at a reasonable cost.

Another important contribution of our work relates to spatio-temporal modelling. The variation over time of the spatial process of helminthiasis risk has so far been neglected. In Chapter 6, we modelled space-time interaction of soil-transmitted helminthiases risk, rather than modelling a temporal trend as a time covariate. The latter assumes that the risk varies uniformly across time, while the former assumes that the spatial process changes over time. Spatio-temporal models were fitted within the SPDE/INLA framework.

7.1.2 Contribution in helminthiases data collection

The paucity of readily available georeferenced epidemiological survey data has long delayed the mapping of helminthiases risk. Thus, it is important to compile, record, regularly update, and share all available data informing about helminthiases prevalence. The Global Neglected Tropical Diseases (GNTD) database has been created within the CONTRAST project (2006–2010), a multidisciplinary alliance for optimizing schistosomiasis control and transmission surveillance in sub-Saharan Africa (Hürlimann et al., 2011; Saarnak et al., 2013). The database is an open-access platform which primarily offered georeferenced schistosomiasis survey prevalence data obtained through a systematic review of published and grey literature. Our work has extended the GNTD to soil-transmitted helminth in Latin America, where data were sorely lacking. In Chapter 4, we reported the results of a systematic review of survey prevalence data pertaining on the three major soil-transmitted helminths *Ascaris lumbricoides*, *Trichuris trichiura* and hookworm in the thirteen countries of South America. Prevalence data were retrieved for 6948 locations by screening 4085 scientific papers. The data are freely available after simple registration on the GNTD website (www.gntd.org). The data retrieved from our literature review enabled us to model the geographical distribution and estimate the risk of ascariasis, trichuriasis and

hookworm infection for the entire South American continent (Chapter 4), as well as for Bolivia (Chapter 5), and, after an update follow up of literature in November 2013, for Brazil (Chapter 6). We additionally updated GNTD with *Schistosoma mansoni* survey data from Côte d'Ivoire and analysed the risk of *S. mansoni* in the country (Chapter 2).

Appart from historical data, we also analysed schistosomiasis data from a recent Ivorian national cross-sectional survey (2012–2013), conducted under the leadership of Dr. Giovanna Raso (Chapter 3). More than 5000 children were examined for multiple parasites in 92 schools. Survey sites were selected following a lattice plus close pairs design (Diggle and Lophaven, 2006), which aimed to optimise spatial coverage at national level.

7.1.3 Contribution in helminthiases epidemiology and implication for control interventions

Our work contributes to the activities of the national control programmes by providing high resolution risk estimates of soil-transmitted helminth infections in South America (Chapter 4, 5, and 6), as well as of schistosomiasis in Côte d'Ivoire (Chapter 2, and 3).

Our geostatistical meta-analysis at continental level of the risk of the three major soil-transmitted helminth species was motivated by the lack of knowledge of their geographical distributions and absence of reliable information in Latin America (Chapter 4). The systematic literature review of survey data highlighted important gaps with regard to soil-transmitted helminths. Our model-based estimates offer important baseline information by suggesting high risk areas where climatic conditions are suitable for the infections. We could identify countries with population-adjusted risk above the 20% threshold advocated by the World Health Organization (WHO) for preventive chemotherapy interventions. Thus, we suggest that national surveys should be prioritised in countries such as French Guiana, Suriname, Venezuela, Bolivia, Guyana, Colombia, Ecuador, Paraguay, Peru, and Chile. We also analysed soil-transmitted helminth risk at national level in Bolivia, where poverty is the highest in South America. Our study indicated that almost half of the population is infected with at least one of the three common soil-transmitted helminths. Although this estimation should be interpreted carefully due to the poor data coverage, it nevertheless confirms that soil-transmitted helminthiasis is of public health concern and that Bolivia is a country where interventions should be prioritised.

An important outcome of our work is that we provided model-based evidence of soil-transmitted helminthiasis risk reduction over time. In South America, we showed that the risk of *A. lumbricoides*, *T. trichiura*, and hookworm has decreased from 2005 onwards.

Motivated by our finding that the risk of soil-transmitted helminth has decreased in South America, we investigated more thoroughly the spatio-temporal dynamic in Brazil (Chapter 6). While the risk of *A. lumbricoides* and hookworm has significantly reduced across time divided into four time periods since 1995, we observed no difference between the period 1995–1999 and from 2010 onwards regarding *T. trichiura* infections. This result might reflect the lack of efficacy of current drugs against trichuriasis (Keiser and Utzinger, 2008) and draws further attention to the need of new chemical components for sustainable control of the infections (Keiser and Utzinger, 2010). Of most importance, our analysis contributes with estimates of the changes in the geographical patterns of risk over time. Thus, our maps inform about the spatial evolution of infection risks. In the light of more recent data and socioeconomic risk factors that were taken into account, the estimated risk turned out lower than suggested by the continental analysis. We believe that our estimation of a soil-transmitted helminthiasis risk of 6.01% reflects better the current Brazilian situation. Therefore, we provided a map highlighting municipalities where the risk was estimated above the cut-off of 20% set up by the WHO for preventive chemotherapy interventions. The recent re-implementation of control measures in Brazil should benefit from these findings for better spatial targeting of their interventions.

Our main contribution to schistosomiasis control consists in estimating the overall schistosomiasis risk based on data from a 2012–2013 national cross-sectional survey in Côte d'Ivoire. We studied the spatial distributions of the two species *S. haematobium* and *S. mansoni*, including co-infection. We provided a map clearly delineating sub-prefectures where the school-aged children risk is estimated above 10%. We believe that this map can guide praziquantel distribution that has recently been implemented by the Ivorian control programme. Furthermore, the geographical patterns of the two infections do not overlap in space. This may reflect the different water preferences of the snails involved in transmission (Steinmann et al., 2006). Importantly, this finding suggests that co-infection with both *Schistosoma* species is rare in Côte d'Ivoire and a potential synergic effect on health is not of concern.

Our work also contributes to better understanding of the factors related to helminthiases in different settings. Variable selection procedures implemented throughout this thesis selected the most important predictors for each infection risk. Generally, those predictors confirmed that warm and humid conditions favour transmission. However, variations of temperature and precipitation have shown their importance in explaining broad scale spatial distribution. Variables such as high isothermality, low range of temperature and

low precipitation seasonality, when selected in the models, were mostly positively associated with soil-transmitted helminthiases risk. Although we could explain that extreme climatic conditions might wash out parasite eggs from the soil or affect their development, those proxies also characterise tropical and subtropical climate, which are known to be particularly suitable for transmission. Another important result was the estimation of a higher risk of trichuriasis in urban setting, which was directly suggested by the influence of variables like human influence index and rurality measure in Brazil, and indirectly by the population density in high risk areas in South America.

7.2 Limitations

Helminthiases are diseases of poverty. Socioeconomic development, access to clean and safe water, as well as sanitation are important factors that must be drastically improved if we want to move towards a significant reduction of the infections burden. Those factors are subject to high disparities between and within countries. To date, we do not have a consistent socioeconomic index able to capture the risk across the different countries of South America. Thus, our continental analysis of soil-transmitted helminthiases risk reflects only the climatic suitabilities of the infections.

Schistosomiasis is known to be a focal disease highly dependent on the presence of the intermediate host snails in the surroundings. In Côte d'Ivoire, the national survey was designed with a sampling approach optimised for spatial modelling. However, small-scale variations in the risk surface were difficult to capture. This raises the difficulties in capturing fine-scale heterogeneities with surveys geolocated at village rather than household level. Additional predictors, such as the presence of the snails or socioeconomic data at higher spatial resolution, could improve estimates of small-scale risk variations.

The analysis of historical data brings a couple of limitations. Our ability to reliably estimate the risk highly depends on the quality of the data on which the models are built. Helminthiases diagnostic techniques vary in sensitivity and specificity (Bergquist et al., 2009). Furthermore, disease risk is age-dependent (Anderson and May, 1985; Raso et al., 2007). Heterogeneities regarding the diagnostic techniques utilised and age categories of the population examined can lead to bias in model-based predictions. Reported historical data often lacks of information regarding age of the survey participants or are aggregated over different age classes, making difficult the estimation of age-specific risks. Surveys may also not report diagnostic methods and sampling efforts. Data are therefore not comparable due to different diagnostic errors or different age groups. Efforts to overcome these issues

within the modelling approach have been undertaken in the past. For example, country-wide age-group alignment factors have been estimated to adjust population-based estimates of *S. mansoni* and *S. haematobium* (Schur et al., 2011a, 2013) and diagnostic performances were incorporated in the modelling of *S. japonicum* risk (Wang et al., 2008). However, diagnostic performances depend on the intensity of infections and age-curve prevalence might be shifted in places where control measures have been conducted. Thus, both age and diagnostic adjustments are expected to vary in space. In Latin America, the data scarcity makes the estimation of age-alignment factors difficult. Immigration-death models can be incorporated into a geostatistical framework and used to age-align the risk across surveys. We are currently working on the development of such models. They are unfortunately not yet available at the time of writing up this thesis.

Poor spatial coverage of the data also affects prediction. A good spatial design is of most importance to reliably estimate the risk in areas where prevalence has not been assessed. Historical data, mainly collected for other purposes than spatial modelling, rarely approach what we could qualify as an optimal design for the study of the geographical distribution. Risk prediction in areas where data are sparse is driven by the relationship between observations and socio-environmental factors established in other areas. Hence, estimation of the risk in areas which are not supported by data should always be interpreted with caution.

Our models were built under the stationary and isotropy assumptions, which postulate the underlying spatial process to be only a function of the distance between locations, irrespective of location and direction. For malaria, it has been shown that the relation between predictors and disease risk varies between different ecological zones (Gosoniu et al., 2009). Non-stationary models have been considered to model infection risks of hookworm (Raso et al., 2006a) and *S. mansoni* (Beck-Wörner et al., 2007) in Côte d'Ivoire. Despite that these models showed the spatial correlation only marginally depends on the locations, a non-stationary spatial process might improve prediction when infection risk is modelled over larger geographical areas. Ecological variation, health systems disparities, and uneven control efforts are among the factors that introduce spatial correlation at local scale, violating the stationary assumption of standard geostatistical models.

The dilemma in modelling historical data often is the inclusion of data that may compromise the precision of the estimates, or the use of a smaller and more homogeneous dataset that may reduce prediction accuracy due to a loss of important information.

7.3 Estimates' comparison

Prior to our work, knowledge of the geographical distribution of soil-transmitted helminths in Latin America was largely unknown. National risk estimates were derived from aggregation of surveys and failed to account for spatial variation (De Silva et al., 2003). Therefore, they cannot be compared to those obtained in this thesis. Geostatistical models, as implemented in our work, provide high resolution estimates (e. g. pixel level) that can be aggregated at any geographic area of interest. Hence, they are important for control programmes, which require estimates at the administrative level at which interventions are carried out.

The 2010 global disease burden study (GDB) (Murray et al., 2013) quantified the worldwide disease burden in disability-adjusted life years (DALYs) for 291 diseases, including helminthiases. The contribution of soil-transmitted helminth infections in the global DALY has been based on estimation of the global numbers of infection recently released by Pullan et al. (2014). In their work, they used a combination of model-based geostatistics for sub-Saharan Africa and analyses based on aggregated data at administrative regions, ignoring small scale variation, for all other regions. We provide a summary of the national estimates from our study and the one by Pullan et al. (2014) in Table 7.1. Main discrepancies concern the estimation of hookworm infection risk, while the overall estimated soil-transmitted helminthiasis risk differs for Ecuador and Suriname. As explained above, these differences show the importance of capturing small-scale variation in disease burden estimation.

Our work provides risk estimations at different scales. For soil-transmitted helminths, we have modelled the risk at continental level (i. e. Latin America), as well as at country level (i. e. Bolivia and Brazil). The question of scale of mapping is an important one that warrants discussion. Socioeconomic inequalities and disparate control efforts between countries are important factors to consider while modelling several countries together. Lack of census in some countries and limited access to detailed socioeconomic information make the task of constructing socioeconomic indices comparable across countries rather difficult (Colston and Saboyá, 2013). Hence, we believe that estimates from national analyses such as the one for Brazil where socioeconomic data were taken into account are more accurate.

Another important comparison to mention concerns the difference between analyses of historical and contemporary data. In Côte d'Ivoire, our analysis of recent schistosomiasis data estimated a lower risk across the country compared to what suggested the analysis of historical data (Chapter 2 and Schur et al. (2013)). We assume that historical data have

a tendency to oversample high endemic areas as research naturally drives data collection in places where infections are known to be of public health concern. This issue is known in literature as preferential sampling and might lead to an overestimation of the risk when such data are utilised. Furthermore, analyses of historical data may not reflect current disease situation, especially in regions where recent data are lacking and interventions are on-going. Hence, good quality data are required to estimate the current situation and to follow its evolution. In this light, it is necessary that endemic countries implement national surveys. The data we used for analysing the overall schistosomiasis risk in Côte d'Ivoire were coming from a national survey, where children have been screened for multiple parasites, including malaria and soil-transmitted helminths. In a cost-effective perspective, it is worth remembering that control programmes should also explore the possibilities to integrate their actions with other programmes.

7.4 Extension of the work

Our work on helminthiases risk estimation aims naturally to be extended to other geographical areas and diseases. Our modelling of soil-transmitted helminthiases risk for South America is a first step towards a "fully geostatistical-based" world map similar to the one released for malaria (Hay et al., 2009). Coupled with risk estimates from China (Lai et al., 2013) and on-going work in Africa and Asia, we will provide a strong basis for updating DALY estimation attributed to soil-transmitted helminths.

An important task that has to be undertaken is the construction of socioeconomic measures that are comparable across countries and are able to capture the spatial variation in the risk of helminthiases. Further research is needed to identify good predictors that could be in the form of a single poverty index, a multidimensional measure (Alkire and Foster, 2011), or more specific proxies related to water and sanitation.

A rigorous approach is also required to age-align estimates across locations in order to obtain age-adjusted risk estimates. We are currently developing immigration-death models to be incorporated within a Bayesian geostatistical framework. These models will improve the estimation of specific age categories.

Table 7.1: Comparison of our geostatistical model-based estimates (Analysis A) with estimates coming from an analysis based on aggregated data at administrative level (Analysis B), in Latin America.

	Estimated risk (%)							
	***Ascaris lumbricoides* infection**		***Trichuris trichiura* infection**		**Hookworm infection**		**Soil-transmitted helminth infection**	
Country	**Analysis A**	**Analysis B**	**Analysis A**	**Analysis B**	**Analysis A**	**Analysis B**	**Analysis A**	**Analysis B**
Argentina	9.5 (6.9; 13.8)	1-10	8.5 (6.3; 11.6)	1–10	8.3 (6.6; 10.3)	1–10	18.9 (16.4; 23.0)	10–20
Bolivia	24.5 (18.4; 29.8)	1–10	18.9 (14.5; 24.9)	1–10	19.3 (15.2; 23.5)	10–20	37.8 (33.2; 41.6)	20–50
Brazil	14.3 (13.4; 16.2)	10–20	10.1 (8.8; 11.3)	1–10	12.3 (10.8; 14.1)	1-10	25.5 (24.2; 27.1)	20–50
Chile	16.8 (13.2; 19.9)	10–20	11.7 (8.4; 15.4)	1–10	7.2 (5.5; 8.8)	< 1	24.7 (21.5; 27.9)	20–50
Colombia	19.7 (16.9; 24.0)	20–50	17.8 (14.4; 21.7)	10–20	14.5 (10.3; 17.5)	10–20	33.4 (30.7; 37.6)	20–50
Ecuador	18.5 (14.9; 24.9)	20–50	11.1 (8.4; 15.1)	20–50	12.3 (8.1; 15.7)	1–10	28.1 (23.8; 33.3)	> 50
French Guiana	25.1 (15.3; 33.0)	10–20	15.1 (7.4; 29.9)	1-10	36.6 (28.8; 44.3)	1–10	46.2 (38.4; 56.1)	20–50
Guyana	23.6 (16.4; 31.6)	10–20	11.1 (5.3; 16.7)	1–10	25.0 (19.5; 29.6)	1–10	37.4 (31.9; 42.9)	20–50
Paraguay	8.3 (5.3; 15.5)	10–20	7.1 (2.6; 21.4)	1–10	18.3 (12.1; 28.0)	1–10	26.7 (20.1; 34.7)	20–50
Peru	17.3 (13.4; 20.6)	10–20	10.8 (8.4; 13.6)	10–20	8.7 (6.1; 11.3)	1–10	25.0 (21.3; 29.0)	20–50
Suriname	23.5 (15.9; 34.8)	1–10	8.3 (3.5; 16.6)	< 1	30.0 (20.3; 41.3)	1–10	40.1 (30.2; 51.1)	1–10
Uruguay	9.9 (6.9; 16.1)	1–10	10.1 (7.3; 15.1)	1–10	6.4 (4.2; 9.2)	1–10	18.8 (14.7; 24.7)	20–50
Venezuela	21.7 (17.9; 25.4)	20–50	27.3 (25.0; 29.8)	20–50	12.2 (9.8; 15.6)	1–10	39.4 (36.3; 42.5)	> 50

Analysis A: estimates presented in this thesis (Chammartin et al., 2013b)
Analysis B: Pullan et al. (2014)
Discrepancies between the two analyses are highlighted in red.

7.5 Concluding remark

Today, progress towards the Millennium Development Goals (MDG) for reducing poverty and improving the lives of deprived population is uneven. Although some countries already reached several objectives, other still have a long way to go and must step up their efforts to reach the target. Fortunately, it has now been decided to support the development of Sustainable Development Goals (SDG) to pursue this ambitious programme beyond 2015 (UN, 2012). WHO has set new milestones for eliminating morbidity due to soil-transmitted helminths by 2020 and reported that only 200 million school-aged children among the 600 million in need received treatment in 2010 (WHO, 2012). Regarding schistosomiasis, the way towards elimination progresses and it is thought that transmission has been interrupted in 19 countries (WHO, 2013b). The disease reference group on helminth infections (DRG4), established by the special programme for research and training in tropical diseases (TDR), recently evaluated research and challenges related to the fight against helminthiases. They concluded that high priority should be placed in the augmentation of the use and application of mathematical models to aid monitoring and evaluation, surveillance, elimination efforts, design of sampling protocols and monitoring of intervention efficacy regarding soil-transmitted helminthiasis and schistosomiasis.

It is in this context that we focussed our work on soil-transmitted helminth infections in South America, where the limited information on the geographical distribution and infection risk have hampered adequate control measures, as well as on schistosomiasis in Côte d'Ivoire, a country where implementation of interventions suffered from a decade of political instabilities. The road towards elimination has to pass by a proper monitoring and evaluation of interventions. Estimation of helminthiases risk is part of a dynamic process and requires updates as long as new data become available and control initiatives moves forward. To conclude, we will continue our efforts to further develop the statistical methodology with regards to the estimation of the burden attributed to helminthiases. We believe that our work provides useful baseline information for control programmes, as well as an important benchmark upon which further estimates could be derived.

Bibliography

Aagaard-Hansen J, Mwanga JR, and Bruun B (2009). Social science perspectives on schistosomiasis control in Africa: past trends and future directions. *Parasitology*, 136(13):1747–1758.

Albonico M, De Carneri I, Di Matteo L, Ghiglietti R, Toscano P, Uledi MK, and Savioli L (1993). Intestinal parasitic infections of urban and rural children on Pemba Island: implications for control. *Annals of Tropical Medicine and Parasitology*, 87(6):579–583.

Alkire S and Foster J (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, 95(7):476–487.

Anderson RM and May RM (1985). Helminth infections of humans: mathematical models, population dynamics, and control. *Advances in Parasitology*, 24:1–101.

Anderson RM, Hollingsworth TD, Truscott J, and Brooker S (2012). Optimisation of mass chemotherapy to control soil-transmitted helminth infection. *The Lancet*, 379(9813):289–290.

Appleton CC (1978). Review of literature on abiotic factors influencing the distribution and life cycles of bilharziasis intermediate host snails. *Malacological Review*, 11:1–25.

Banerjee S, Gelfand AE, and Carlin BP (2003). *Hierarchical modeling and analysis for spatial data*. Chapman & Hall/CRC.

Banerjee S, Gelfand AE, Finley AO, and Sang H (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):825–848.

Barbieri MM and Berger JO (2004). Optimal predictive model selection. *The Annals of Statistics*, 32(3):870–897.

Basset D, Gaumerais H, and Basset-Pougnet A (1986). Intestinal parasitoses in children of an Indian community of Bolivian Altiplano. *Bulletin de la Société de Pathologie Exotique et de ses filiales*, 79(2):237–246.

Batjes NH (2006). ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (ver. 1.1). Report 2006/02. Wageningen: ISRIC-World Soil Information.

Beaumier CM, Bethony J, Loukas A, Loblack M, Diemert D, Bottazzi ME, and Hotez PJ (2012). Human hookworm vaccine, a vaccine to prevent intestinal blood loss and transmission caused by the hookworm *Necator americanus*. *Journal of Vaccines & Vaccination*, 3(4):85.

Beck-Wörner C, Raso G, Vounatsou P, N'Goran EK, Rigo G, Parlow E, and Utzinger J (2007). Bayesian spatial risk prediction of *Schistosoma mansoni* infection in western Côte d'Ivoire using a remotely-sensed digital elevation model. *The American Journal of Tropical Medicine and Hygiene*, 76(5):956–963.

Belizán JM, Cafferata ML, Belizán M, and Althabe F (2007). Health inequality in Latin America. *The Lancet*, 370(9599):1599–1600.

Benefice E, Monroy SL, Jiménez S, and López R (2006). Nutritional status of Amerindian children from the Beni River (lowland Bolivia) as related to environmental, maternal and dietary factors. *Public Health Nutrition*, 9(3):327–335.

Bergquist R, Johansen MV, and Utzinger J (2009). Diagnostic dilemmas in helminthology: what tools to use and when? *Trends in Parasitology*, 25(4):151–156.

Best N, Richardson S, and Thomson A (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*, 14(1):35–59.

Bethony J, Brooker S, Albonico M, Geiger SM, Loukas A, Diemert D, and Hotez PJ (2006). Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. *The Lancet*, 367(9521):1521–1532.

Bitran R, Martorell B, Escobar L, Munoz R, and Glassman A (2009). Controlling and eliminating neglected diseases in Latin America and the Caribbean. *Health Affairs*, 28(6):1707–1719.

Bogoch II, Andrews JR, Speich B, Utzinger J, Ame SM, Ali SM, and Keiser J (2013). Short report: Mobile phone microscopy for the diagnosis of soil-transmitted helminth infections: A proof-of-concept study. *The American Journal of Tropical Medicine and Hygiene*, 88(4):626–629.

Bonfoh B, Raso G, Koné I, Dao D, Girardin O, Cissé G, Zinsstag J, Utzinger J, and Tanner M (2011). Research in a war zone. *Nature*, 474(7353):569–571.

Booth M, Bundy DAP, Albonico M, Chwaya HM, Alawi KS, and Savioli L (1998). Associations among multiple geohelminth species infections in schoolchildren from Pemba Island. *Parasitology*, 116(01):85–93.

Booth M, Vounatsou P, N'Goran EK, Tanner M, and Utzinger J (2003). The influence of sampling effort and the performance of the Kato-Katz technique in diagnosing *Schistosoma mansoni* and hookworm co-infections in rural Côte d'Ivoire. *Parasitology*, 127(6):525–531.

Brooker S (2002). Schistosomes, snails and satellites. *Acta Tropica*, 82(2):207–214.

Brooker S (2007). Spatial epidemiology of human schistosomiasis in Africa: risk models, transmission dynamics and control. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 101(1):1–8.

Brooker S and Clements ACA (2009). Spatial heterogeneity of parasite co-infection: determinants and geostatistical prediction at regional scales. *International Journal for Parasitology*, 39(5):591–597.

Brooker S, Beasley M, Ndinaromtan M, Madjiouroum EM, Baboguel M, Djenguinabe E, Hay SI, and Bundy DAP (2002). Use of remote sensing and a geographical information system in a national helminth control programme in Chad. *Bulletin of the World Health Organization*, 80(10):783–789.

Brooker S, Whawell S, Kabatereine NB, Fenwick A, and Anderson RM (2004a). Evaluating the epidemiological impact of national control programmes for helminths. *Trends in Parasitology*, 20(11):537–545.

Brooker S, Kabatereine NB, Tukahebwa EM, and Kazibwe F (2004b). Spatial analysis of the distribution of intestinal nematode infections in Uganda. *Epidemiology & Infection*, 132(6):1065–1071.

Brooker S, Clements ACA, and Bundy DAP (2006). Global epidemiology, ecology and control of soil-transmitted helminth infections. *Advances in Parasitology*, 62:221–261.

Brooker S, Kabatereine NB, O GJ, Stothard JP, and Utzinger J (2009a). Rapid mapping of schistosomiasis and other neglected tropical diseases in the context of integrated control programmes in Africa. *Parasitology*, 136(13):1707–1718.

Brooker S, Kabatereine NB, Smith JL, Mupfasoni D, Mwanje MT, Ndayishimiye O, Lwambo NJS, Mbotha D, Karanja P, Mwandawiro C, Muchiri E, Clements ACA, and Snow RW (2009b). An updated atlas of human helminth infections: the example of East Africa. *International Journal of Health Geographics*, 8(1):42.

Brooker S, Hotez PJ, and Bundy DAP (2010). The global atlas of helminth infection: mapping the way forward in neglected tropical disease control. *PLoS Neglected Tropical Diseases*, 4(7):e779.

Brown DS (1994). *Freshwater snails of Africa and their medical importance, 2nd edition*. Taylor & Francis/CRC.

Brown HW (1927). Studies on the rate of development and viability of the eggs of *Ascaris lumbricoides* and *Trichuris trichiura* under field conditions. *The Journal of Parasitology*, 14(1):1–15.

Bundy DAP and Cooper ES (1989). *Trichuris* and trichuriasis in humans. *Advances in Parasitology*, 28:107–173.

Cameletti M, Lindgren F, Simpson D, and Rue H (2013). Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *Advances in Statistical Analysis*, 97(2):109–131.

Cancrini G, Bartoloni A, Paradisi F, and Nunez LE (1989). Parasitological observations on three Bolivian localities including rural communities, cities and institutions. *Annals of Tropical Medicine and Parasitology*, 83(6):591–594.

Cancrini G, Bartoloni A, Zaffaroni E, Guglielmetti P, Gamboa N, Nicoletti A, and Genchi C (1998). Seroprevalence of *Toxocara canis*-IgG antibodies in two rural Bolivian communities. *Parassitologia*, 40:473–476.

Chammartin F, Hürlimann E, Raso G, N'Goran EK, Utzinger J, and Vounatsou P (2013a). Statistical methodological issues in mapping historical schistosomiasis survey data. *Acta Tropica*, 128(2):345–352.

Chammartin F, Scholte RGC, Guimarães LH, Tanner M, Utzinger J, and Vounatsou P (2013b). Soil-transmitted helminth infection in South America: a systematic review and geostatistical meta-analysis. *The Lancet Infectious Diseases*, 13(6):507–518.

Chammartin F, Scholte RGC, Malone JB, Bavia ME, Nieto P, Utzinger J, and Vounatsou P (2013c). Modelling the geographical distribution of soil-transmitted helminth infections in Bolivia. *Parasites & Vectors*, 6(1):1–14.

Chan MS, Guyatt HL, Bundy DAP, and Medley GF (1994). The development and validation of an age-structured model for the evaluation of disease control strategies for intestinal helminths. *Parasitology*, 109:389–389.

CIESIN, FAO and CIAT (2005). Center for International Earth Science Information Network Columbia University, UN Food and Agriculture Programme, Centro Internacional

de Agricultura Tropical. Gridded population of the world: future estimates (GPWFE). Palisades NY, USA: Socioeconomic Data and Applications Center (SEDAC) Columbia University.

Clements ACA, Lwambo NJS, Blair L, Nyandindi U, Kaatano G, Kinung'hi S, Webster JP, Fenwick A, and Brooker S (2006). Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. *Tropical Medicine & International Health*, 11(4):490–503.

Clements ACA, Bosqué-Oliva E, Sacko M, Landouré A, Dembelé R, Traoré M, Coulibaly G, Gabrielli AF, Fenwick A, and Brooker S (2009a). A comparative study of the spatial distribution of schistosomiasis in Mali in 1984–1989 and 2004–2006. *PLoS Neglected Tropical Diseases*, 3(5):e431.

Clements ACA, Firth S, Dembelé R, Garba A, Touré S, Sacko M, Landouré A, Bosqué-Oliva E, Barnett AG, Brooker S, and Fenwick A (2009b). Use of Bayesian geostatistical prediction to estimate local variations in *Schistosoma haematobium* infection in western Africa. *Bulletin of the World Health Organization*, 87(12):921–929.

Clements ACA, Deville MA, Ndayishimiye O, Brooker S, and Fenwick A (2010). Spatial co-distribution of neglected tropical diseases in the East African Great Lakes region: revisiting the justification for integrated control. *Tropical Medicine & International Health*, 15(2):198–207.

Clennon JA, King CH, Muchiri EM, Kariuki HC, Ouma JH, Mungai P, and Kitron U (2004). Spatial patterns of urinary schistosmiasis infection in a highly endemic area of coastal Kenya. *American Journal of Tropical Medicine and Hygiene*, 70(4):443–448.

Colley DG, Bustinduy AL, E SW, and King CH (2014). Human schistosomiasis. *The Lancet*, 383(4):2253–64.

Colston J and Saboyá M (2013). Soil-transmitted helminthiasis in Latin America and the Caribbean: modelling the determinants, prevalence, population at risk and costs of control at sub-national level. *Geospatial Health*, 7(2):321–340.

Coulibaly JT, Knopp S, N'Guessan NA, Silué KD, Fürst T, Lohourignon LK, Brou JK, N'Gbesso YK, Vounatsou P, N'Goran EK, and Utzinger J (2011). Accuracy of urine circulating cathodic antigen (CCA) test for *Schistosoma mansoni* diagnosis in different settings of Côte d'Ivoire. *PLoS Neglected Tropical Diseases*, 5(11):e1384.

Craig MH, Sharp BL, Mabaso MLH, and Kleinschmidt I (2007). Developing a spatial-statistical model and map of historical malaria prevalence in Botswana using a staged variable selection procedure. *International Journal of Health Geographics*, 6(1):44.

Cressie N (1990). The origins of kriging. *Mathematical Geology*, 22(3):239–252.

Cressie N and Johannesson G (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):209–226.

Dacombe RJ, Crampin AC, Floyd S, Randall A, Ndhlovu R, Bickle Q, and Fine PEM (2007). Time delays between patient and laboratory selectively affect accuracy of helminth diagnosis. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 101(2):140–145.

De Moira AP, Fulford AJC, Kabatereine NB, Kazibwe F, Ouma JH, Dunne DW, and Booth M (2007). Microgeographical and tribal variations in water contact and *Schistosoma mansoni* exposure within a Ugandan fishing community. *Tropical Medicine & International Health*, 12(6):724–735.

De Silva NR and Hall A (2010). Using the prevalence of individual species of intestinal nematode worms to estimate the combined prevalence of any species. *PLoS Neglected Tropical Diseases*, 4(4):e655.

De Silva NR, Brooker S, Hotez PJ, Montresor A, Engels D, and Savioli L (2003). Soil-transmitted helminth infections: updating the global picture. *Trends in Parasitology*, 19(12):547–551.

De Vlas SJ and Gryseels B (1992). Underestimation of *Schistosoma mansoni* prevalences. *Parasitology Today*, 8(8):274–277.

Dellaportas P, Forster JJ, and Ntzoufras I (2002). On Bayesian model and variable selection using MCMC. *Statistics and Computing*, 12(1):27–36.

Diggle PJ and Lophaven SÃ (2006). Bayesian geostatistical design. *Scandinavian Journal of Statistics*, 33(1):53–64.

Diggle PJ, Tawn JA, and Moyeed RA (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):299–350.

Diggle PJ, Moyeed R, Rowlingson B, and Thomson M (2002). Childhood malaria in the Gambia: a case-study in model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 51(4):493–506.

Doumenge JP, Mott KE, Cheung C, Villenave D, Chapuis O, Perrin MF, and Reaud-Thomas G (1987). *Atlas of the global distribution of schistosomiasis*. Presses Universitaires de Bordeaux.

Engels D, Chitsulo L, Montresor A, and Savioli L (2002). The global epidemiological situation of schistosomiasis and new approaches to control and research. *Acta Tropica*, 82(2):139–146.

Enk MJ, Lima ACL, Drummond SC, Schall VT, and Coelho PMZ (2008). The effect of the number of stool samples on the observed prevalence and the infection intensity with *Schistosoma mansoni* among a population in an area of low transmission. *Acta Tropica*, 108(2):222–228.

Esteban JG, Flores A, Aguirre C, Strauss W, Angles R, and Mas-Coma S (1997). Presence of very high prevalence and intensity of infection with *Fasciola hepatica* among Aymara children from the northern Bolivian Altiplano. *Acta Tropica*, 66(1):1–14.

Ezeamama AE, Friedman JF, Olveda RM, Acosta LP, Kurtis JD, Mor V, and McGarvey ST (2005). Functional significance of low-intensity polyparasite helminth infections in anemia. *Journal of Infectious Diseases*, 192(12):2160–2170.

Fenwick A and Savioli L (2011). Schistosomiasis elimination. *The Lancet Infectious Diseases*, 11:346.

Flores A, Esteban JG, Angles R, and Mas-Coma S (2001). Soil-transmitted helminth infections at very high altitude in Bolivia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 95(3):272–277.

French MD, Rollinson D, Basáñez MG, Mgeni AF, Khamis I, and Stothard JR (2007). School-based control of urinary schistosomiasis on Zanzibar, Tanzania: monitoring micro-haematuria with reagent strips as a rapid urological assessment. *Journal of Pediatric Urology*, 3(5):364–368.

Furrer R, Genton MG, and Nychka D (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3).

Gabrielli AF, Montresor A, Nicholls RS, and Ault SK (2013). Progress towards the control and elimination of neglected tropical diseases in Brazil. *Jornal de Pediatria*, 89(3):215–216.

Garba A, Lamine MS, Barkiré N, Djibo A, Sofo B, Gouvras AN, Labbo R, Sebangou H, Webster JP, Fenwick A, and Utzinger J (2012). Efficacy and safety of two closely spaced doses of praziquantel against *Schistosoma haematobium* and *S. mansoni* and re-infection patterns in school-aged children in Niger. *Acta Tropica*, 128(2):334–344.

Gelfand AE and Smith AFM (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409.

Gelman A, Van Dyk DA, Huang Z, and Boscardin JW (2008). Using redundant parameterizations to fit hierarchical models. *Journal of Computational and Graphical Statistics*, 17(1):95–122.

Gemperli A (2003). Development of spatial statistical methods for modelling point-referenced spatial data in malaria epidemiology. University of Basel (PhD thesis).

Gemperli A, Vounatsou P, Kleinschmidt I, Bagayoko M, Lengeler C, and Smith T (2004). Spatial patterns of infant mortality in Mali: the effect of malaria endemicity. *American Journal of Epidemiology*, 159(1):64–72.

Gemperli A, Sogoba N, Fondjo E, Mabaso M, Bagayoko M, Briët OJ, Anderegg D, Liebe J, Smith T, and Vounatsou P (2006). Mapping malaria transmission in West and Central Africa. *Tropical Medicine & International Health*, 11(7):1032–1046.

George EI and McCulloch RE (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.

Giardina F, Gosoniu L, Konate L, Diouf MB, Perry R, Gaye O, Faye O, and Vounatsou P (2012). Estimating the burden of malaria in Senegal: Bayesian zero-inflated binomial geostatistical modeling of the MIS 2008 data. *PLoS One*, 7(3):e32625.

Goodman D, Haji HJ, Bickle QD, Stoltzfus RJ, Tielsch JM, Ramsan M, Savioli L, and Albonico M (2007). A comparison of methods for detecting the eggs of *Ascaris*, *Trichuris*, and hookworm in infant stool, and the epidemiology of infection in Zanzibari infants. *American Journal of Tropical Medicine and Hygiene*, 76(4):725–731.

Gosoniu L, Vounatsou P, Sogoba N, Maire N, and Smith T (2009). Mapping malaria risk in West Africa using a Bayesian nonparametric non-stationary model. *Computational Statistics & Data Analysis*, 53(9):3358–3371.

Gunawardena GSA, Karunaweera ND, and Ismail MM (2004). Wet-days: are they better indicators of *Ascaris* infection levels? *Journal of Helminthology*, 78(4):305–310.

Gyapong JO, Gyapong M, Yellu N, Anakwah K, Amofah G, Bockarie M, and Adjei S (2010). Integration of control of neglected tropical diseases into health-care systems: challenges and opportunities. *The Lancet*, 375(9709):160–165.

Hall A, Hewitt G, Tuffrey V, and De Silva NR (2008). A review and meta-analysis of the impact of intestinal worms on child growth and nutrition. *Maternal & Child Nutrition*, 4(s1):118–236.

Hastings WK (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, Noor AM, Kabaria CW, Manh BH, Elyazar IRF, Brooker S, Smith DL, Moyeed RA, and Snow RW (2009). A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Medicine*, 6(3):e1000048.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, and Jarvis A (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15):1965–1978.

Hodges MH, Soares Magalhães RJ, Paye J, Koroma JB, Sonnie M, Clements ACA, and Zhang Y (2012). Combined spatial prediction of schistosomiasis and soil-transmitted helminthiasis in Sierra Leone: a tool for integrated disease control. *PLoS Neglected Tropical Diseases*, 6(6):e1694.

Holford TR and Hardy RJ (1976). A stochastic model for the analysis of age-specific prevalence curves in schistosomiasis. *Journal of Chronic Diseases*, 29(7):445–458.

Hotez PJ (2008). Hookworm and poverty. *Annals of the New York Academy of Sciences*, 1136:38–44.

Hotez PJ (2011). A handful of "antipoverty" vaccines exist for neglected diseases, but the world's poorest billion people need more. *Health Affairs*, 30(6):1080–1087.

Hotez PJ and Pritchard DI (1995). Hookworm infection. *Scientific American*, 272(6):68–75.

Hotez PJ, Raff S, Fenwick A, Richards Jr F, and Molyneux DH (2007). Recent progress in integrated neglected tropical disease control. *Trends in Parasitology*, 23(11):511–514.

Hotez PJ, Brindley PJ, Bethony JM, King CH, Pearce EJ, and Jacobson J (2008a). Helminth infections: the great neglected tropical diseases. *The Journal of Clinical Investigation*, 118(4):1311.

Hotez PJ, Molyneux DH, Fenwick A, Savioli L, and Takeuchi T (2008b). A global fund to fight neglected tropical diseases: is the G8 Hokkaido Toyako 2008 summit ready? *PLoS Neglected Tropical Diseases*, 2(3):e220.

Hotez PJ, Fenwick A, Savioli L, and Molyneux DH (2009). Rescuing the bottom billion through control of neglected tropical diseases. *The Lancet*, 373(9674):1570–1575.

Hürlimann E, Schur N, Boutsika K, Stensgaard AS, Laserna de Himpsl M, Ziegelbauer K, Laizer N, Camenzind L, Di Pasquale A, Ekpo UF, Simoonga C, Mushinge G, Saarnak CFL, Utzinger J, Kristensen TK, and Vounatsou P (2011). Toward an open-access global database for mapping, control, and surveillance of neglected tropical diseases. *PLoS Neglected Tropical Diseases*, 5(12):e1404.

IADB, PAHO and SABIN (2011). Inter-American Development Bank, Pan American Health Organization, Sabin Vaccine Institute. A call to action: addressing soil-transmitted helminths in Latin America and the Caribbean. http://new.paho.org/hq/index.php?option=com_docman&task=doc_download&gid=13723Itemid=, accessed January 31, 2013.

Ishwaran H and Rao JS (2005). Spike and slab variable selection: frequentist and Bayesian strategies. *Annals of Statistics*, 33(2):730–773.

Johnson ME, Moore LM, and Ylvisaker D (1990). Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, 26(2):131–148.

Karagiannis-Voules DA, Scholte RGC, Guimarães LH, Utzinger J, and Vounatsou P (2013). Bayesian geostatistical modeling of leishmaniasis incidence in Brazil. *PLoS Neglected Tropical Diseases*, 7(5):e2213.

Karagiannis-Voules DA, Odermatt P, Biedermann P, Khieu V, Schär F, Muth S, Utzinger J, and Vounatsou P (2014). Geostatistical modelling of soil-transmitted helminth infection in Cambodia: do socioeconomic factors improve predictions? *Acta tropica*, in press. doi: 10.1016/j.actatropica.2014.09.001.

Keiser J and Utzinger J (2008). Efficacy of current drugs against soil-transmitted helminth infections. *JAMA: the Journal of the American Medical Association*, 299(16):1937–1948.

Keiser J and Utzinger J (2010). The drugs we have and the drugs we need against major helminth infections. *Advances in Parasitology*, 73:197–230.

Kloos H (1995). Human behavior, health education and schistosomiasis control: a review. *Social Science & Medicine*, 40(11):1497–1511.

Knopp S, Mgeni AF, Khamis IS, Steinmann P, Stothard JR, Rollinson D, Marti H, and Utzinger J (2008). Diagnosis of soil-transmitted helminths in the era of preventive chemotherapy: effect of multiple stool sampling and use of different diagnostic techniques. *PLoS Neglected Tropical Diseases*, 2(11):e331.

Knopp S, Rinaldi L, Khamis IS, Stothard JR, Rollinson D, Maurelli MP, Steinmann P, Marti H, Cringoli G, and Utzinger J (2009). A single FLOTAC is more sensitive than triplicate Kato–Katz for the diagnosis of low-intensity soil-transmitted helminth infections. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 103(4):347–354.

Knopp S, Steinmann P, Keiser J, and Utzinger J (2012). Nematode infections: soil-transmitted helminths and *Trichinella*. *Infectious Disease Clinics of North America*, 26(2):341–358.

Knorr-Held L (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, 19(17-18):2555–2567.

Koukounari A, Donnelly CA, Sacko M, Keita AD, Landouré A, Dembelé R, Bosqué-Oliva E, Gabrielli AF, Gouvras A, Traoré M, Fenwick A, and Webster JP (2010). The impact of single versus mixed schistosome species infections on liver, spleen and bladder morbidity within Malian children pre-and post-praziquantel treatment. *BMC Infectious Diseases*, 10(1):227.

Krauth SJ, Coulibaly JT, Knopp S, Traoré M, N'Goran EK, and Utzinger J (2012). An in-depth analysis of a piece of shit: distribution of *Schistosoma mansoni* and hookworm eggs in human stool. *PLoS Neglected Tropical Diseases*, 6(12):e1969.

Kuo L and Mallick B (1998). Variable selection for regression models. *Sankhyā: The Indian Journal of Statistics, Series B*, 60:65–81.

Lai YS, Zhou XN, Utzinger J, and Vounatsou P (2013). Bayesian geostatistical modelling of soil-transmitted helminth survey data in the People's Republic of China. *Parasites & Vectors*, 6(1):1–16.

Lengeler C, Utzinger J, and Tanner M (2002). Questionnaires for rapid screening of schistosomiasis in sub-Saharan Africa. *Bulletin of the World Health Organization*, 80(3):235–242.

Levin SA (1992). The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology*, 73(6):1943–1967.

Lindgren F, Rue H, and Lindström J (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Loukas A, Bethony J, Brooker S, and Hotez PJ (2006). Hookworm vaccines: past, present, and future. *The Lancet Infectious Diseases*, 6(11):733–741.

Machault V, Vignolles C, Borchi F, Vounatsou P, Briolant S, Lacaux JP, and Rogier C (2011). The use of remotely sensed environmental data in the study of malaria. *Geospatial Health*, 5(2):151–168.

Malone JB (2005). Biology-based mapping of vector-borne parasites by geographic information systems and remote sensing. *Parassitologia*, 47(1):27–50.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, and Teller E (1953). Equation

of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21:1087.

Minsiterio da Saúde-Brasil (a). Programa integrado de equistossomose da Fundação Oswaldo Cruz (PIDE). http://pide.cpqrr.fiocruz.br/index.php?pagina=3&PHPSESSID=f11d881b7822379b490aeb4665c4c608, accessed January 31, 2013.

Minsiterio da Saúde-Brasil (b). Equistossomose mansônica. Document CID 10: B65. http://portal.saude.gov.br/portal/arquivos/pdf/gve_7ed_web_atua_equistissomose_mansonica.pdf, accessed January 31, 2013.

Moher D, Liberati A, Tetzlaff J, and Altman DG (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, 151(4):264–269.

Mollinedo S and Prieto C (2006). El enteroparasitismo en Bolivia. La Paz: Ministerio de Salud y Deportes.

Molyneux DH, Hotez PJ, and Fenwick A (2005). "Rapid-impact interventions": how a policy of integrated control for Africa's neglected tropical diseases could benefit the poor. *PLoS Medicine*, 2(11):e336.

Montresor A, Crompton DWT, Hall A, Bundy DAP, and Savioli L (1998). Guidelines for the evaluation of soil-transmitted helminthiasis and schistosomiasis at community level (WHO/CDS/SIP/98.1). Geneva: World Health Organization.

Montresor A, Odermatt P, Muth S, Iwata F, Raja'a YA, Assis AM, Zulkifli A, Kabatereine NB, Fenwick A, Al-Awaidy S, Allenk H, Engelsk D, and Savioli K (2005). The WHO dose pole for the administration of praziquantel is also accurate in non-African populations. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 99(1):78–81.

Mudenda NB, Malone JB, Kearney MT, Mischler PD, Nieto P, McCarroll JC, and Vounatsou P (2012). Modelling the ecological niche of hookworm in Brazil based on climate. *Geospatial Health*, 6(3):S111–S123.

Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, Ezzati M, Shibuya K, Salomon JA, Abdalla S, et al. (2013). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the global burden of disease study 2010. *The Lancet*, 380(9859):2197–2223.

O'Hara RB and Sillanpää MJ (2009). A review of Bayesian variable selection methods: what, how and which. *Bayesian Analysis*, 4(1):85–117.

Oliveira MAG (2010). Sources of Brazil's counter-hegemony. *Revista Brasileira de Politica Internacional*, 53(2):125–141.

Otto G (1929). A study of the moisture requirements of the eggs of the horse, the dog, human and pig ascarids. *American Journal of Epidemiology*, 10(2):497–520.

PAHO and WHO (2009). 61*st* session of the regional committee: elimination of neglected diseases and other poverty-related infections, resolution CD49.R19. Washington, DC: Pan American Health Organization. http://new.paho.org/hq/dmdocuments/2009/CD49.R19%20(Eng.).pdf, accessed January 31, 2013.

Pullan RL and Brooker S (2012). The global limits and population at risk of soil-transmitted helminth infections in 2010. *Parasites & Vectors*, 5(2):81.

Pullan RL, Sturrock HJW, Soares Magalhães RJ, Clements ACA, and Brooker SJ (2012). Spatial parasite ecology and epidemiology: a review of methods and applications. *Parasitology*, 139(14):1870–1887.

Pullan RL, Smith JL, Jasrasaria R, and Brooker SJ (2014). Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. *Parasites & Vectors*, 7(1):37.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raftery AE and Lewis S (1992). How many iterations in the Gibbs sampler. *Bayesian Statistics*, 4(2):763–773.

Raso G, Matthys B, N'Goran EK, Tanner M, Vounatsou P, and Utzinger J (2005). Spatial risk prediction and mapping of *Schistosoma mansoni* infections among schoolchildren living in western Côte d'Ivoire. *Parasitology*, 131(01):97–108.

Raso G, Vounatsou P, Gosoniu L, Tanner M, N'Goran EK, and Utzinger J (2006a). Risk factors and spatial patterns of hookworm infection among schoolchildren in a rural area of western Côte d'Ivoire. *International Journal for Parasitology*, 36(2):201–210.

Raso G, Vounatsou P, Singer BH, N'Goran EK, Tanner M, and Utzinger J (2006b). An integrated approach for risk profiling and spatial prediction of *Schistosoma mansoni*–hookworm coinfection. *Proceedings of the National Academy of Sciences*, 103(18):6934–6939.

Raso G, Vounatsou P, McManus DP, N'Goran EK, and Utzinger J (2007). A Bayesian approach to estimate the age-specific prevalence of *Schistosoma mansoni* and implications for schistosomiasis control. *International Journal for Parasitology*, 37(13):1491–1500.

Riedel N, Vounatsou P, Miller JM, Gosoniu L, Chizema-Kawesha E, Mukonka V, and Steketee RW (2010). Geographical patterns and predictors of malaria risk in Zambia: Bayesian geostatistical modelling of the 2006 Zambia national malaria indicator survey (ZMIS). *Malaria Journal*, 9:37.

Robinson E, Picon D, Sturrock HJ, Sabasio A, Lado M, Kolaczinski J, and Brooker S (2009). The performance of haematuria reagent strips for the rapid mapping of urinary schistosomiasis: field experience from Southern Sudan. *Tropical Medicine & International Health*, 14(12):1484–1487.

Rollinson D (2009). A wake up call for urinary schistosomiasis: reconciling research effort with public health importance. *Parasitology*, 136(12):1593–1610.

Rollinson D, Knopp S, Levitz S, Stothard JR, Tchuem Tchuenté LA, Garba A, Mohammed KA, Schur N, Person B, Colley DG, and J U (2013). Time to set the agenda for schistosomiasis elimination. *Acta Tropica*, 128(2):423–440.

Rue H, Martino S, and Chopin N (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Saarnak CFL, Utzinger J, and Kristensen TK (2013). Collection, verification, sharing and dissemination of data: the CONTRAST experience. *Acta Tropica*, 128(2):407–411.

Scheipl F, Fahrmeir L, and Kneib T (2012). Spike-and-slab priors for function selection in structured additive regression models. *Journal of the American Statistical Association*, 107(500):1518–1532.

Schneider MC, Aguilera XP, Da Silva Junior JB, Ault SK, Najera P, Martinez J, Requejo R, Nicholls RS, Yadon Z, Silva JC, Leanes LF, and Periago MR (2011). Elimination of neglected diseases in Latin America and the Caribbean: a mapping of selected diseases. *PLoS Neglected Tropical Diseases*, 5(2):e964.

Scholte RGC, Schur N, Bavia ME, Carvalho EM, Chammartin F, Utzinger J, and Vounatsou P (2013). Spatial analysis and risk mapping of soil-transmitted helminths in Brazil, using Bayesian geostatistical models. *Geospatial Health*, 8(1):97–110.

Schur N (2011). Geostatistical modelling of schistosomiasis transmission in Africa. University of Basel (PhD thesis).

Schur N, Hürlimann E, Garba A, Traoré MS, Ndir O, Ratard RC, Tchuem Tchuenté LA,

Kristensen TK, Utzinger J, and Vounatsou P (2011a). Geostatistical model-based estimates of schistosomiasis prevalence among individuals aged ≤20 years in West Africa. *PLoS Neglected Tropical Diseases*, 5(6):e1194.

Schur N, Utzinger J, and Vounatsou P (2011b). Modelling age-heterogeneous *Schistosoma haematobium* and *S. mansoni* survey data via alignment factors. *Parasites & Vectors*, 4(1):142.

Schur N, Vounatsou P, and Utzinger J (2012). Determining treatment needs at different spatial scales using geostatistical model-based risk estimates of schistosomiasis. *PLoS Neglected Tropical Diseases*, 6(9):e1773.

Schur N, Hürlimann E, Stensgaard AS, Chimfwembe K, Mushinge G, Simoonga C, Kabatereine NB, Kristensen TK, Utzinger J, and Vounatsou P (2013). Spatially explicit *Schistosoma* infection risk in eastern Africa using Bayesian geostatistical modelling. *Acta Tropica*, 128(2):365–377.

Simoonga C, Utzinger J, Brooker S, Vounatsou P, Appleton CC, Stensgaard AS, Olsen A, and Kristensen TK (2009). Remote sensing, geographical information system and spatial analysis for schistosomiasis epidemiology and ecology in Africa. *Parasitology*, 136(13):1683–1693.

Smith G, Schad GA, and Warren KS (1990). The ecology of the free-living stages: a reappraisal. In *Hookworm disease-current status and new directions*. Taylor & Francis.

Smits HL (2009). Prospects for the control of neglected tropical diseases by mass drug administration. *Expert Review of Anti-Infective Therapy*, 7(1):37–56.

Soares Magalhães RJ and Clements ACA (2011). Mapping the risk of anaemia in preschool-age children: the contribution of malnutrition, malaria, and helminth infections in West Africa. *PLoS Medicine*, 8(6):e1000438.

Soares Magalhães RJ, Clements ACA, Patil AP, Gething PW, and Brooker S (2011). The applications of model-based geostatistics in helminth epidemiology and control. *Advances in Parasitology*, 74:267–296.

Southgate VR (1997). Schistosomiasis in the Senegal River Basin: before and after the construction of the dams at Diama, Senegal and Manantali, Mali and future prospects. *Journal of Helminthology*, 71(2):125–132.

Speich B, Knopp S, Mohammed KA, Khamis IS, Rinaldi L, Cringoli G, Rollinson D, and Utzinger J (2010). Comparative cost assessment of the Kato-Katz and FLOTAC

techniques for soil-transmitted helminth diagnosis in epidemiological surveys. *Parasites & Vectors*, 3(71).

Spiegelhalter DJ, Best NG, Carlin BP, and Van Der Linde A (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Spindler LA (1929). The relation of moisture to the distribution of human *Trichuris* and *Ascaris*. *American Journal of Epidemiology*, 10(2):476–496.

Steinmann P, Keiser J, Bos R, Tanner M, and Utzinger J (2006). Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *The Lancet Infectious Diseases*, 6(7):411–425.

Steinmann P, Du ZW, Wang LB, Wang XZ, Jiang JY, Li LH, Marti H, Zhou XN, and Utzinger J (2008). Extensive multiparasitism in a village of Yunnan province, People's Republic of China, revealed by a suite of diagnostic methods. *American Journal of Tropical Medicine and Hygiene*, 78(5):760–769.

Stensgaard AS, Utzinger J, Vounatsou P, Hürlimann E, Schur N, Saarnak CF, Simoonga C, Mubita P, Kabatereine NB, Tchuem Tchuenté LA, Rahbek C, and Kristensen TK (2013). Large-scale determinants of intestinal schistosomiasis and intermediate host snail distribution across Africa: does climate matter? *Acta Tropica*, 128(2):378–390.

Strunz EC, Addiss DG, Stocks ME, Ogden S, Utzinger J, and Freeman MC (2014). Water, sanitation, hygiene, and soil-transmitted helminth infection: a systematic review and meta-analysis. *PLoS Medicine*, 11(3):e1001620.

Sturrock HJ, Gething PW, Ashton RA, Kolaczinski JH, Kabatereine NB, and Brooker S (2011). Planning schistosomiasis control: investigation of alternative sampling strategies for *Schistosoma mansoni* to target mass drug administration of praziquantel in East Africa. *International Health*, 3(3):165–175.

Tanner SN (2005). A population in transition: health, culture change, and intestinal parasitism among the Tsimane' of lowland Bolivia. Northwestern University (PhD thesis).

Taylor P and Makura O (1985). Prevalence and distribution of schistosomiasis in Zimbabwe. *Annals of Tropical Medicine and Parasitology*, 79(3):287–299.

Tchuem Tchuenté LA and N'Goran EK (2009). Schistosomiasis and soil-transmitted helminthiasis control in Cameroon and Côte d'Ivoire: implementing control on a limited budget. *Parasitology*, 136(13):1739.

UN (2012). The future we want. 66*th* session UN General Assembly. Resolution adopted on July 27 2012 (A/RES/66/288). New York: United Nations.

Utzinger J, Booth M, N'Goran EK, Müller I, Tanner M, and Lengeler C (2001). Relative contribution of day-to-day and intra-specimen variation in faecal egg counts of *Schistosoma mansoni* before and after treatment with praziquantel. *Parasitology*, 122(05):537–544.

Utzinger J, Raso G, Brooker S, De Savigny D, Tanner M, Ornbjerg N, Singer BH, and N'Goran EK (2009). Schistosomiasis and neglected tropical diseases: towards integrated and sustainable control and a word of caution. *Parasitology*, 136(13):1859.

Utzinger J, Tanner M, and Keiser J (2010). Acts for schistosomiasis: do they act? *The Lancet Infectious Diseases*, 10(9):579–581.

Utzinger J, N'Goran EK, Caffrey CR, and Keiser J (2011). From innovation to application: social–ecological context, diagnostics, drugs and integrated control of schistosomiasis. *Acta Tropica*, 120:S121–S137.

Van der Werf MJ, De Vlas SJ, Brooker S, Looman CWN, Nagelkerke NJD, Habbema JDF, and Engels D (2003). Quantification of clinical morbidity associated with schistosome infection in sub-Saharan Africa. *Acta Tropica*, 86(2):125–139.

Ver Hoef JM, Cressie N, Fisher RN, and Case TJ (2001). Uncertainty and spatial linear models for ecological data. In *Spatial uncertainty in ecology: implications for remote sensing and GIS applications*. Springer-Verlag.

Vounatsou P, Raso G, Tanner M, N'Goran EK, and Utzinger J (2009). Bayesian geostatistical modelling for mapping schistosomiasis transmission. *Parasitology*, 136(13):1695–1705.

Wang XH, Zhou XN, Vounatsou P, Chen Z, Utzinger J, Yang K, Steinmann P, and Wu XH (2008). Bayesian spatio-temporal modeling of *Schistosoma japonicum* prevalence data in the absence of a diagnostic "gold" standard. *PLoS Neglected Tropical Diseases*, 2(6):e250.

Warren KS (1973). Regulation of the prevalence and intensity of schistosomiasis in man: immunology or ecology? *Journal of Infectious Diseases*, 127(5):595–609.

WHO (2002a). Helminth control in school-age children: a guide for managers of control programmes 1*st* edition. Geneva: World Health Organization.

WHO (2002b). Prevention and control of schistosomiasis and soil-transmitted helminthiasis. Technical report series: 912. Geneva: World Health Organization.

WHO (2006). Preventive chemotherapy in human helminthiasis: coordinated use of anthelminthic drugs in control interventions: a manual for health professionals and programme managers. Geneva: World Health Organization.

WHO (2009). Elimination of schistosomiasis from low transmission areas. Report of a WHO informal consultation (WHO/HTM/NTD/PCT/2009.2). Geneva: World Health Organization.

WHO (2010). Working to overcome the global impact of neglected tropical diseases. First WHO report on neglected tropical diseases (WHO/HTM/NTD/2010.1). Geneva: World Health Organization.

WHO (2011a). Helminth control in school-age children: a guide for managers of control programmes. 2*nd* edition. Geneva: World Health Organization.

WHO (2011b). Soil-transmitted helminthiases: estimates of the number of children needing preventive chemotherapy and number treated, 2009. *Weekly Epidemiological Record*, 86:257–268.

WHO (2012). Eliminating soil-transmitted helminthiases as a public health problem in children. Progress report 2001–2010 and strategic plan 2011–2020 (WHO/HTM/NTD/PCT/2012.4). Geneva: World Health Organization.

WHO (2013a). Assessing the efficacy of anthelminthic drugs against schistosomiasis and soil-transmitted helminthiases (WHO/HTM/NTD/PCT/2013.4). Geneva: World Health Organization.

WHO (2013b). Schistosomiasis. Progress report 2001–2010 and strategic plan 2011–2020 (WHO/HTM/NTD/PCT/2013.2). Geneva: World Health Organization.

WHO (2014). Schistosomiasis: number of people receiving preventiv chemotherapy in 2012. *Weekly Epidemiological Record*, 89:21–28.

WHO and UNICEF (2006). Joint monitoring programme for water supply and sanitation. Core questions on drinking water and sanitation for household surveys. Geneva, New York: World Health Organization, United Nations Children's Fund.

Woolhouse MEJ (1998). Patterns in parasite epidemiology: the peak shift. *Parasitology Today*, 14(10):428–434.

Xiao SH, Utzinger J, Tanner M, Keiser J, and Xue J (2013). Advances with the Chinese anthelminthic drug tribendimidine in clinical trials and laboratory investigations. *Acta Tropica*, 126(2):115–126.

Yadav A (2003). Development and survival of *Ascaris lumbricoides* eggs under the high rainfall and humid conditions prevailing in Meghalaya, India. *Proceedings of the Zoological Society (Calcutta)*, 56:109–112.

Yajima A, Gabrielli AF, Montresor A, and Engels D (2011). Moderate and high endemicity of schistosomiasis is a predictor of the endemicity of soil-transmitted helminthiasis: a systematic review. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 105(2):68–73.

Yapi RB, Hürlimann E, Houngbedji CA, Ndri PB, Silué KD, Soro G, Kouamé FN, Vounatsou P, Fürst T, N'Goran EK, Utzinger J, and Raso G (2014). Infection and co-infecton of helminths and *Plasmodium* among school children in Côte d'Ivoire: results from a national cross-sectional survey. *PLoS Neglected Tropical Diseases*, 8(6):e2913.

Yekutiel P (1981). Lessons from the big eradication campaigns. *World Health Forum*, 2(4):465–490.

Zhang Z and Jiang Q (2011). Schistosomiasis elimination. *The Lancet Infectious Diseases*, 11:345.

Ziegelbauer K, Speich B, Mäusezahl D, Bos R, Keiser J, and Utzinger J (2012). Effect of sanitation on soil-transmitted helminth infection: systematic review and meta-analysis. *PLoS Medicine*, 9(1):e1001162.

Curriculum vitae

Frédérique Chammartin

Date and place of birth:	August 21, 1975 in Billens, Switzerland
Nationality:	Swiss
E-mail address:	f.chammartin@unibas.ch, fredcham@yahoo.com

Education

2014	PhD in Epidemiology at the Swiss Tropical and Public Health Institute/University of Basel, Basel, Switzerland Thesis: Geostatistical model-based predictions of helminthiases risk to assist control interventions (supervision: PD Dr. P. Vounatsou and Prof. Dr. J. Utzinger)
2010	Master of Science (MSc) in Statistics at the University of Neuchâtel, Neuchâtel, Switzerland Thesis: Modelling the geographical distribution of *Schistosoma mansoni*, hookworm and *Plasmodium falciparum* co-infection risk in a rural area of western Côte d'Ivoire (supervision: PD Dr. P. Vounatsou and Prof. Dr. C. Stărică)
2001	Diploma (Dipl Fed) in Pharmacy at the University of Lausanne, Lausanne, Switzerland

Professional Activities

Since March 2014	Research fellow in Biostatistics at Swiss Tropical and Public Health Institute/University of Basel, Basel, Switzerland
2011–2014	PhD fellow in Epidemiology at the Swiss Tropical and Public Health Institute/University of Basel, Basel, Switzerland
2008–2010	Pharmacist itinerant
2007	Pharmacist in charge at Calcutta Rescue, Kolkata, India
2004–2007	Pharmacist in charge at Pharmacie Capitole Bullet, Estavayer-le-Lac, Switzerland
2002–2003	Pharmacist at Pharmacie du Câro, Bulle, Switzerland

Achievements and Conferences

2013	Bayesian Inference for Latent Gaussian Models with Applications, Reykjavík, Iceland, workshop attendance
2013	Spatial Statistics 2013, Columbus, Ohio, USA, poster presentation on "Geostatistical modelling with variable selection: a comparison of different approaches"
2012	Barcelona Biomed Conference on Bayesian methods in Biostatistics and Bioinformatics, poster presentation on "Soil-transmitted helminth infection in South America: a systematic review and geostatistical meta-analysis"

Publications

Chammartin, F, Scholte, RGC, Guimarães, LH, Tanner, M, Utzinger, J, and Vounatsou, P (2013). **Soil-transmitted helminth infection in South America: a systematic review and geostatistical meta-analysis.** *The Lancet Infectious Diseases*, 13(6): 507-518.

Chammartin, F, Hürlimann, E, Raso, G, N'Goran, EK, Utzinger, J, and Vounatsou, P (2013). **Statistical methodological issues in mapping historical schistosomiasis survey data.** *Acta Tropica*, 128(2): 345–352.

Chammartin, F, Scholte, RGC, Malone, JB, Bavia, ME, Nieto, P, Utzinger, J, and Vounatsou, P (2013). **Modelling the geographical distribution of soil-transmitted helminth infections in Bolivia.** *Parasites & Vectors*, 6: 152.

Murto, C, **Chammartin, F**, Schwarz, K, da Costa, LMM, Kaplan, C, and Heukelbach, J (2013). **Patterns of migration and risks associated with leprosy among migrants in Maranhão, Brazil.** *PLoS Neglected Tropical Diseases*, 7(9): e2422.

Scholte, RGC, Schur, N, Bavia, ME, Carvalho, EM, **Chammartin, F**, Utzinger, J, and Vounatsou, P (2013). **Spatial analysis and risk mapping of soil-transmitted helminth infections in Brazil, using Bayesian geostatistical models.** *Geospatial Health*, 8(1): 97–110.

Scholte, RGC, Gosoniu, L, Malone, JB, **Chammartin, F**, Utzinger, J, and Vounatsou, P (2014). **Predictive risk mapping of schistosomiasis in Brazil using Bayesian geostatistical models.** *Acta Tropica*, 132: 57–63.

Chammartin, F, Guimarães, LH, Scholte, RGC, Bavia, ME, Utzinger, J, and Vounatsou, P (2014). **Spatio-temporal distribution of soil-transmitted helminth infections in Brazil.** *Parasites & Vectors*, 7: 440.

Chammartin, F, Houngbedji CA, Hürlimann E, Yapi RB, Silué KD, Soro G , Kouamé FN, N'Goran EK, Utzinger J, Raso G and Vounatsou P (2014). ***Schistosoma haematobium*–*Schistosoma mansoni* co-distribution in Côte d'Ivoire.** *PLoS Neglected Tropical Diseases*, in press.